

1. Statistica, probabilità.

La statistica, come scienza, si occupa dei metodi per raccogliere, ordinare, riassumere, presentare ed analizzare i dati, anche per trarre valide conclusioni e prendere ragionevoli decisioni sulla base di tali analisi. Essa sta coprendo un ruolo sempre più importante in pressoché tutti i campi del comportamento umano. All'inizio, la statistica è servita solo per descrivere le caratteristiche degli stati (da cui il nome), ma poi la sua influenza ha raggiunto agricoltura, biologia, affari, medicina, fisica, politica, psicologia e numerosi altri campi della scienza. Su fisica essa ha trovato una sua collocazione specifica nello sviluppo della meccanica statistica, nella interpretazione della meccanica quantistica, e i suoi concetti si ritrovano un po' dappertutto nella fisica.

Uno dei concetti di base della statistica è quello di probabilità. La probabilità è lo studio degli esperimenti casuali o non deterministici. Se un dado viene lanciato in aria, è certo che il dado ricadrà, ma non è certo, poniamo, che si presenterà il numero 6 sulla faccia superiore. Tuttavia, supponiamo di ripetere questo esperimento del lancio del dado un gran numero di volte; sia s il numero di volte che si presenta il numero 6 sulla faccia superiore e sia n il numero di lanci eseguiti. E' stato empiricamente osservato che il rapporto

$$f = \frac{s}{n} \quad (1.1)$$

detto frequenza relativa, a lungo andare si stabilizza, cioè tende a un limite per n molto grande. Questa stabilità costituisce la base della teoria delle probabilità.

Sulla base di tale stabilità è allora possibile costruire un modello matematico di un fenomeno non deterministico

(2)

assegnando una probabilità (il valore limite della frequenza relativa) agli eventi connessi con l'esperimento. Naturalmente, l'affidabilità del nostro modello matematico per un dato evento, dipende dalla misura in cui le probabilità assegnate approssimano le effettive frequenze relative.

Storicamente la teoria delle probabilità iniziò con lo studio dei giochi d'azzardo, come la roulette e i giochi di carte. La probabilità p di un evento A era definita nel modo seguente:

"Se A può verificarsi in s modi su di un totale di n modi egualmente verosimili, allora

$$p = P(A) = \frac{s}{n}.$$

esprime la probabilità dell'evento A ".

Questa definizione classica di probabilità, basata sostanzialmente sulle osservazioni delle frequenze relative, pur essendo immediata e intuitiva, è in realtà contraddittoria, in quanto contiene un circolo vizioso. La nozione di "egualmente verosimile", che equivale a "egualmente probabile", viene usata prima di essere definita e entra nella sua stessa definizione.

Il trattamento moderno della teoria delle probabilità è meramente assiomatico. Ciò significa che le probabilità dei nostri eventi possono essere assolutamente arbitrarie, con l'eccezione che esse debbono soddisfare certi assiomi ben precisi. In questo modo è possibile dare risposte chiare e inequivocabili a questioni precise e ben poste. La teoria delle probabilità rimane però una branca in cui, malgrado le sue regole e definizioni precise e rigorose, è facilissimo prendere degli abbagli, anche per degli esperti. La conferma di ciò viene dalla storia. Leibnitz pensava che fosse egualmente

facile ottenere 11 o 12 con una coppia di dadi. Jean le Rond d'Alembert, il grande matematico francese del XVIII secolo, non riusciva a vedere che i risultati ottenibili lanciando una moneta tre volte sono uguali a quelli di tre monete lanciate contemporaneamente e credeva (come molti giocatori persistono nel credere) che dopo una lunga serie di "teste", una "croce" sia più probabile. In base a tale credenza, i giocatori del lotto puntano cifre sempre più alte sui numeri ritardatari, con vantaggio finale dello stato (che tiene il "banco").

La teoria delle probabilità può condurre a risultati che sembrano in contrasto con il senso comune, quasi paradossali. Consideriamo il "paradosso del secondo figlio". Il signor Rossi dice: "Io ho due figli e almeno uno di essi è maschio". Quale è la probabilità che anche l'altro figlio è maschio? Si sarebbe tentati di dire $\frac{1}{2}$, mentre in realtà la probabilità è $\frac{1}{3}$. Infatti le possibilità per i due figli sono, in generale (M: maschio, F: femmina)

M M	} "almeno uno è maschio"
M F	
F M	
F F	

dicendo che almeno un figlio è maschio esclude la possibilità che siano entrambi femmine. Rimangono solo 3 possibilità di cui una sola è quella favorevole. Attenzione però ai tranelli. Supponiamo di lanciare in aria due monete, di cui una viene tenuta nascosta, mentre l'altra viene osservata. Assumiamo che sia testa e ragioniamo: "abbiamo due monete e la mia mostra testa. La probabilità che anche l'altra mostri testa è $\frac{1}{3}$, pertanto scommetto che è croce". A lungo andare perdo, perché nel ragionamento ho commesso un errore, e la vera probabilità che anche l'altra sia

④

testa è $\frac{1}{2}$. L'errore consiste nello specificare quale moneta presenta testa (la mia, quella che vedo). E' come se il signor Rossi avesse detto "mio figlio maggiore (o più alto, o più grasso) è maschio", questo riduce le possibilità a due sole (MM e MF) e la probabilità che anche l'altro sia maschio è $\frac{1}{2}$.

Questo esempio fa capire che la teoria delle probabilità fornisce delle risposte chiare e inequivocabili a questioni semplici, ma solo quando il procedimento sperimentale usato per ottenere l'evento è definito con precisione. La mancanza di questo requisito è una fonte comune di confusione.

Come abbiamo detto, con la teoria della probabilità si possono ottenere risultati molto strani, verità che appaiono così in contrasto con il senso comune da essere difficilmente credibili anche dopo che ci si è trovati di fronte alla loro dimostrazione. Il paradosso delle date di nascita ne è un esempio dei più puri. Scegliendo a caso 24 persone quale ritenete sia la probabilità che due o più di essi abbiano lo stesso giorno di nascita? (ovvia lo stesso giorno e mese dell'anno, indipendentemente dall'anno di nascita?) A intuito viene ritenuta molto bassa. In realtà è superiore al 50% (53.83%)!

George Gamow, in "Uno, due, tre, ... infinito" dà il seguente semplice procedimento per giungere all'inatteso risultato. La probabilità che i compleanni di due persone qualsiasi non cadano nello stesso giorno è chiaramente $\frac{364}{365}$ (vi è una sola possibilità su 365 che il compleanno di una persona coincida con quello di un'altra). La probabilità che il compleanno di una terza persona differisca da quello delle altre due è $\frac{363}{365}$; per una quarta è $\frac{362}{365}$; e così via fino alla 24^a persona ($\frac{342}{365}$). Otteniamo così 23 frazioni che devono essere moltiplicate tra loro per avere la

probabilità che tutti i 24 compleanni siano differenti.

Il prodotto finale è una frazione che vale circa

$23/50$. Quindi la probabilità di una coincidenza

di almeno due compleanni su 24 risulta circa $27/50$!

In questo calcolo abbiamo ignorato il fatto che alcuni

anni sono bisestili e che i compleanni tendono a

concentrarsi più in certi mesi che in altri. Il primo

fatto tende a diminuire la probabilità, il secondo ad

aumentarla. Nella tabella successiva vengono

mostrati alcuni valori della probabilità di coincidenza

dei compleanni di due persone (almeno) su n .

La certezza assoluta viene raggiunta solo con 366 persone, ma già con 60 persone la probabilità supera il 99%.

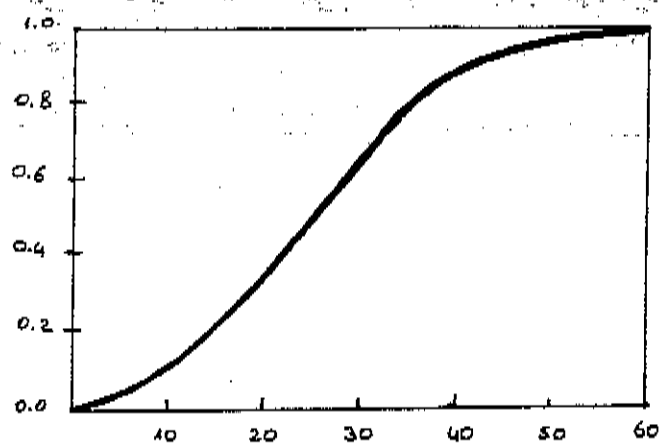
Probabilità (percentuale) p_n che almeno due persone su n abbiano il compleanno lo stesso giorno.

$p_2 =$	0.27 %	$p_{20} =$	41.1 %
$p_3 =$	0.82 %	$p_{25} =$	56.9 %
$p_4 =$	1.63 %	$p_{30} =$	70.6 %
$p_5 =$	2.71 %	$p_{35} =$	81.4 %
$p_6 =$	4.04 %	$p_{40} =$	89.1 %
$p_7 =$	5.62 %	$p_{45} =$	94.1 %
$p_8 =$	7.43 %	$p_{50} =$	97.0 %
$p_9 =$	9.46 %	$p_{55} =$	98.6 %
$p_{10} =$	11.7 %	$p_{60} =$	99.4 %
$:$		$p_{70} =$	99.9 %
$p_{15} =$	25.3 %	$p_{80} =$	99.99 %

$p_{366} = 100. \%$

6

probabilità
di coincidenza



numero di persone

2. Probabilità cumulativa.

Da un punto di vista matematico moderno la probabilità è definita in maniera assiomatica astratta. In questo senso la teoria della probabilità risulta una particolare teoria della misura. In fisica siamo interessati ai risultati di misure fisiche, che possono rappresentare angoli di deflessione, differenze di potenziale, energie, conteggi di decadimenti, ecc. Il risultato è espresso da numeri reali, e possiamo dire che questo individua un punto in uno spazio numerico a 1 o più dimensioni. Questo ci permette di limitarci al caso di probabilità definite su insiemi numerici, in particolare \mathbb{R} .

In effetti è sempre possibile associare ai vari modi di presentarsi di un fenomeno un risultato numerico. Nel lancio di una moneta possono verificarsi due casi, comunemente detti "testa" e "croce". Possiamo allora associare a "testa" il valore '0' e a "croce" il valore '1' (oppure qualsiasi altra coppia di valori). Un evento in questo caso è rappresentato da una variabile numerica ξ che può assumere due valori come possibili risultati della misura (tutti gli altri valori risultano impossibili).

Quello che abbiamo costruito rientra nel concetto più generale di variabile casuale (o random). Una variabile casuale è una associazione tra i risultati di un esperimento (soggetto ad una certa aleatorietà) e un numero reale. Molto spesso la variabile casuale ξ rappresenta proprio il risultato di una misura.

La distribuzione di valori della variabile casuale dipenderà dal modo in cui abbiamo costruito l'associazione e dalla distribuzione degli eventi possibili. Non assumeremo il concetto di probabilità come intuitivo e chiaro. Per esempio l'affermazione che un neutrone ha una probabilità

⑧

0.316 di passare attraverso un certo strato di materiale
vogliamo indicare due cose

- il neutrone si è comportato in questo modo nel 31.6% dei casi su un grande numero di casi.
- crediamo fermamente che in futuro il neutrone si comporterà in maniera simile, e la percentuale sperimentale tenderà a un valore del 31.6% all'aumentare in maniera indefinita del numero dei tentativi.

In linea di principio il range di valori di una variabile casuale è tutto \mathbb{R} . Se però certi valori non vengono mai assunti questi valori vengono associati all'evento impossibile (con probabilità nulla). Sappiamo che i risultati di una misura possono essere rappresentati da quantità variabili con continuità in certi intervalli, oppure da quantità che possono assumere solo valori discretizzati. Nel caso di risultati discreti (ad esempio il conteggio di decadimenti radiativi) possiamo associare ai valori una probabilità finita espressa da un numero compreso tra 0 e 1 (0 è la probabilità degli eventi impossibili, 1 rappresenta la certezza assoluta) identificabile sperimentalmente con la frequenza relativa (1.1) per $n \rightarrow \infty$. Nel caso di variabili continue è privo di senso assegnare a un risultato una probabilità finita. Sperimentalmente possiamo solo dire che il risultato apparterrà, con una certa probabilità, ad un certo intervallo, più o meno ampio.

In generale possiamo dire che associando una variabile casuale con i possibili eventi di un esperimento abbiamo assegnato una funzione di probabilità P per gli intervalli di valori della variabile casuale ξ . Diamo cioè significato alla probabilità

$$P\{x_1 \leq \xi \leq x_2\}$$

intendendo il senso più esteso per la parola "intervallo", aperto, chiuso, limitato, illimitato, o anche degenerare fino a coincidere con un punto isolato quando ha senso parlare di probabilità in un punto (nel caso di variabili discrete).

Naturalmente nulla ci vieta di costruire una associazione tra gli eventi possibili e coppie di variabili casuali, o in generale, n -uple $(\xi_1, \xi_2, \dots, \xi_n)$ di variabili casuali, dando significato a probabilità del tipo:

$$P \{ x_1 < \xi_1 \leq x_2; y_1 \leq \xi_2 < y_2; \dots z_1 \leq \xi_n \leq z_2 \}$$

nel caso $n=1$ parliamo di probabilità e distribuzioni univariate, nel caso $n>1$ le diremo multivariate (bivariate nel caso $n=2$), e gli intervalli (in senso lato) sono da intendersi come intervalli a più dimensioni o pluriintervalli.

Per ora ci limitiamo al caso $n=1$, univariato, per introdurre i concetti. Matematicamente la probabilità P (e gli insiemi su cui è definita) e le sue proprietà sono codificate mediante assiomi e regole ben precise. Il risultato di queste regole ed assiomi si possono riassumere dicendo che tutte le informazioni sulla variabile casuale ξ da un punto di vista probabilistico sono contenute nella funzione

$$F(x) = P \{ \xi \leq x \}$$

(2.1)

detta probabilità cumulativa. Tramite la funzione $F(x)$ definita dalla (2.1) è possibile esprimere altre probabilità. Ad esempio

$$P \{ x_1 < \xi \leq x_2 \} = F(x_2) - F(x_1)$$

$$P\{\xi > x\} = 1 - F(x)$$

La funzione $F(x)$ non è certamente una funzione arbitraria, ma deve soddisfare alcune proprietà derivanti dalla sua definizione e dalla proprietà di P (che abbiamo assunto come intuitive). Dobbiamo avere, in base alla definizione:

i) $F(x)$ è una funzione monotona crescente

$$a \leq b \Rightarrow F(a) \leq F(b) \quad (2.2)$$

$$\text{ii) } F(-\infty) = 0; \quad F(+\infty) = 1 \quad (2.3)$$

iii) $F(x)$ è continua da destra:

$$\lim_{\epsilon \rightarrow 0^+} F(x + \epsilon) = F(x) \quad (2.4)$$

Notiamo che non è detto che sia verificata la continuità completa. La funzione $F(x)$ può presentare dei salti. Se salto viene a rappresentare una probabilità discreta (cioè concentrata in un punto), valutabile tramite la relazione

$$P\{\xi = x_0\} = F(x_0 + 0) - F(x_0 - 0) \quad (2.5)$$

La continuità da destra è una conseguenza della scelta della disuguaglianza in senso lato e non stretto nella definizione (2.1). Tale scelta è arbitraria, e se avessimo scelto la disuguaglianza in senso stretto ($<$ invece di \leq) avremmo avuto la proprietà di continuità da sinistra.

In certe trattazioni viene fatta tale scelta senza che la teoria subisca modifiche sostanziali. Occorre fare attenzione ed essere sempre coerenti con la scelta operata.

In ogni caso avremmo avuto la possibilità di "salti" nella funzione $F(x)$ (2.1). Spesso è possibile esprimere le varie relazioni in maniera indipendente da tale scelta; un esempio è fornito dalla (2.5) che esprime una probabilità finita associata a un valore della variabile casuale ξ .

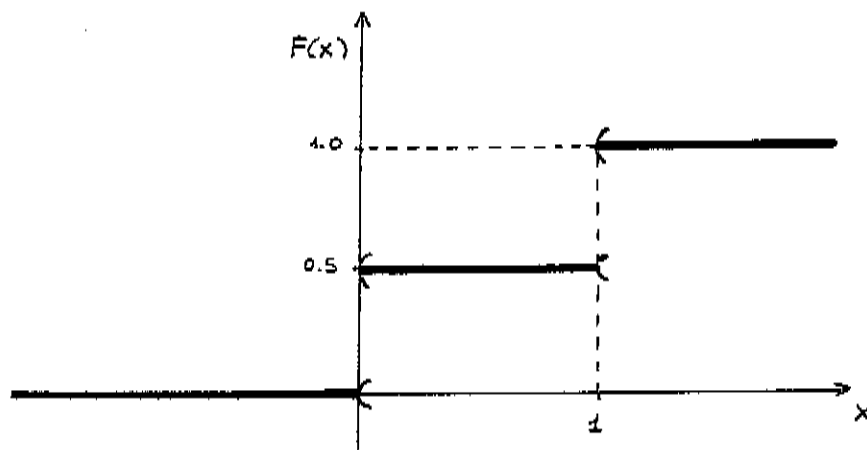
L'importanza della probabilità cumulativa $F(x)$ risiede nel fatto che essa contiene tutte le informazioni sulla funzione di probabilità P . In effetti si può dimostrare che qualsiasi funzione $F(x)$ che verifica le proprietà (i), (ii), (iii) descrive una probabilità P per una variabile casuale. Questo ci permette di dire che le relazioni (2.2), (2.3), (2.4) per una funzione $F(x)$ definiscono una probabilità, e le possiamo assumere come punto di partenza della nostra trattazione.

Vediamo ora di chiarire con qualche esempio le caratteristiche di una probabilità cumulativa $F(x)$.

Esempio 1. Lancio di una moneta.

Come già accennato, ai possibili risultati del lancio di una moneta possiamo associare i valori 0 ("testa") e 1 ("croce"), e possiamo assumere, o verificare, sperimentalmente eseguendo un grande numero di lanci, nel caso di una perfetta equivalenza tra le due facce, che entrambi i risultati abbiano probabilità $\frac{1}{2}$ di verificarsi. Allora possiamo descrivere tutte le possibilità del processo (lancio di una moneta) tramite la probabilità cumulativa (univariata):

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$



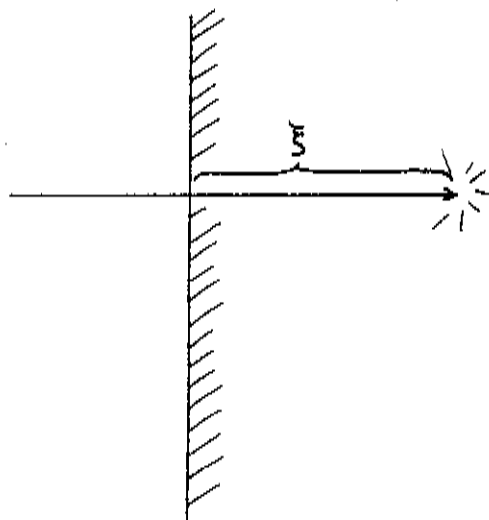
Vediamo che la probabilità cumulativa è rappresentata da una funzione a scalini (continua da destra) e nei punti di discontinuità ($x = 0, 1$) $F(x)$ assume come valore il bordo superiore del gradino.

Esempio 8. Penetrazione di neutroni nella materia.

Un neutrone (o un fotone) entra in un materiale uniforme e percorre una distanza ξ prima di subire una collisione fortemente anelastica, in cui, per semplicità, assumiamo che venga assorbito. È naturale assumere che, almeno per tratti piccoli, la probabilità che tale neutrone venga assorbito (o comunque subisca un urto anelastico) in un tratto di percorso dX sia proporzionale al percorso stesso dX (dX indica un tratto infinitesimo del percorso). Tale probabilità sarà data dal rapporto tra la diminuzione ($-dN$) di particelle nel tratto dX (tale diminuzione rappresenta il numero di particelle che subiscono la collisione) e il numero di particelle stesse N presenti nel tratto dX :

$$-\frac{dN}{N} = \frac{dX}{\lambda} \quad (2.6)$$

dove abbiamo indicato con $1/\lambda$ la costante di proporzionalità. λ , che ha le dimensioni di una lunghezza, dipenderà dal tipo di materiale e dal tipo di particelle usate, e viene detta libero cammino medio. Notiamo che la variazione di numero delle particelle dN è negativo, per cui $-dN$ è positivo. Assumendo che i neutroni entrino nel materiale nella posizione $X = 0$, abbiamo:



(14)

$$N(x) = N_0 e^{-\frac{x}{\lambda}} \quad (2.7)$$

dove N_0 è il numero di particelle entranti nel materiale. $N(x)$

representa il numero di particelle presenti, e che non hanno ancora subito una collisione anelastica, dopo un tratto x .

Il rapporto

$$\frac{N(x)}{N_0} = e^{-\frac{x}{\lambda}} \quad (2.8)$$

individa quindi la probabilità che una particella sia ancora illusa ad una distanza x e che subisca quindi una collisione ad una distanza $\xi > x$:

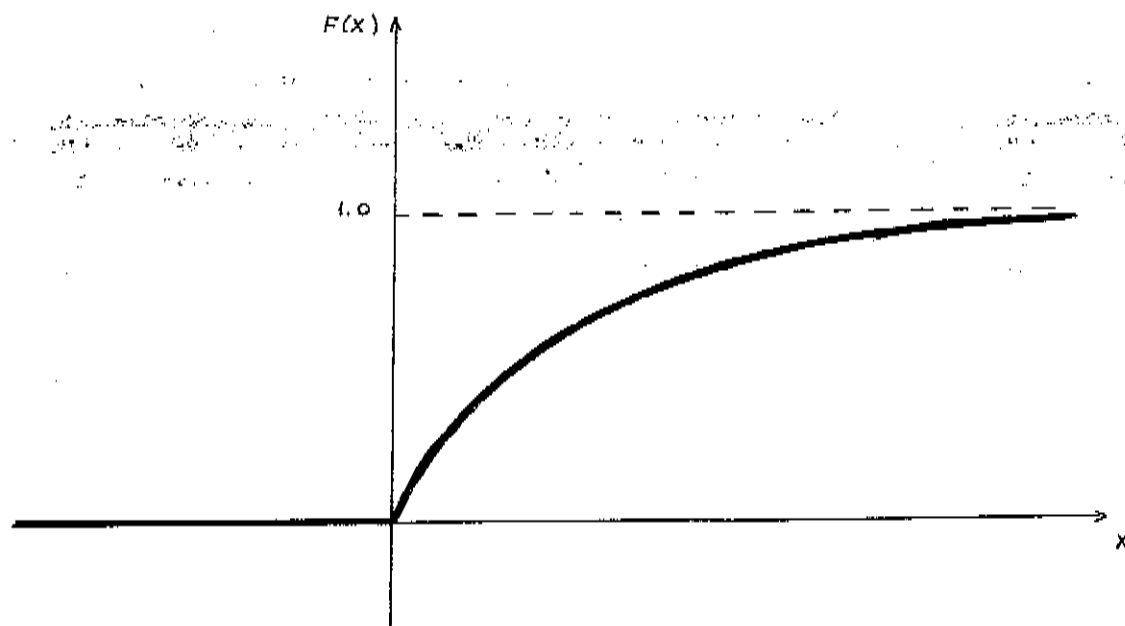
$$P\{\xi > x\} = e^{-\frac{x}{\lambda}}$$

Pertanto:

$$F(x) = P\{\xi \leq x\} = 1 - e^{-\frac{x}{\lambda}} \quad (2.9)$$

Ovviamente, il ragionamento sopra è valido per ξ e $x \geq 0$. Non ha senso parlare di un tratto ξ prima della collisione negativo, per cui:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-x/\lambda} & \text{se } x \geq 0 \end{cases} \quad (2.10)$$



Vediamo che in questo caso la probabilità cumulativa è una funzione continua, derivabile ovunque escluso nell'origine ($x=0$), con derivata continua.

$$f(x) = F'(x) \quad (2.11)$$

Questa derivata viene detta in generale densità di probabilità e si può scrivere

$$F(x) = \int_{-\infty}^x f(x) dx \quad (2.12)$$

con

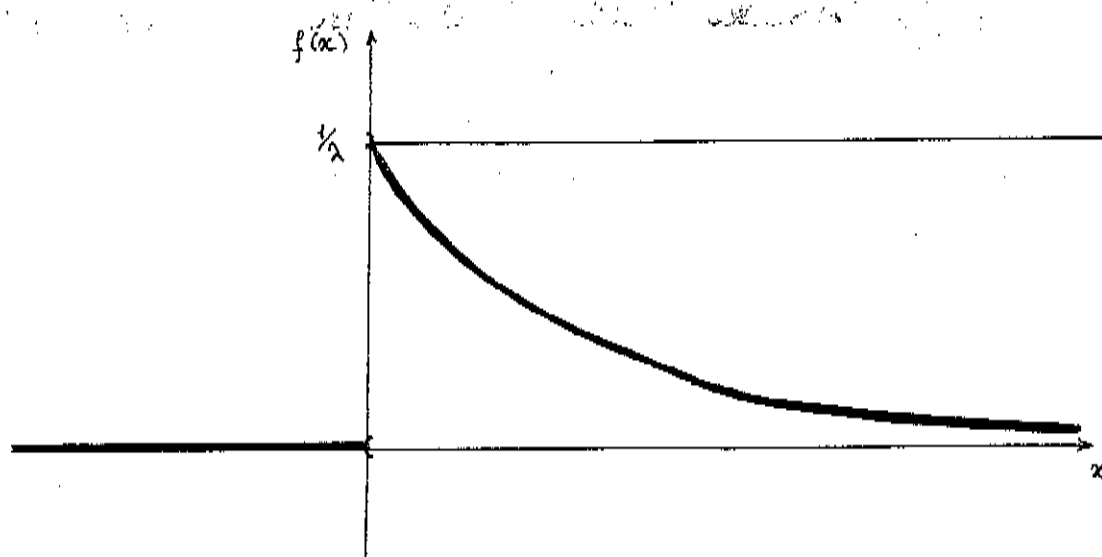
$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & x > 0 \end{cases} \quad (2.13)$$

La funzione $F(x)$ in questo caso risulta una funzione continua, anzi, per amor di precisione, rientra nella classe delle funzioni assolutamente continue, con derivata (quasi ovunque)

$$f(x) \geq 0$$

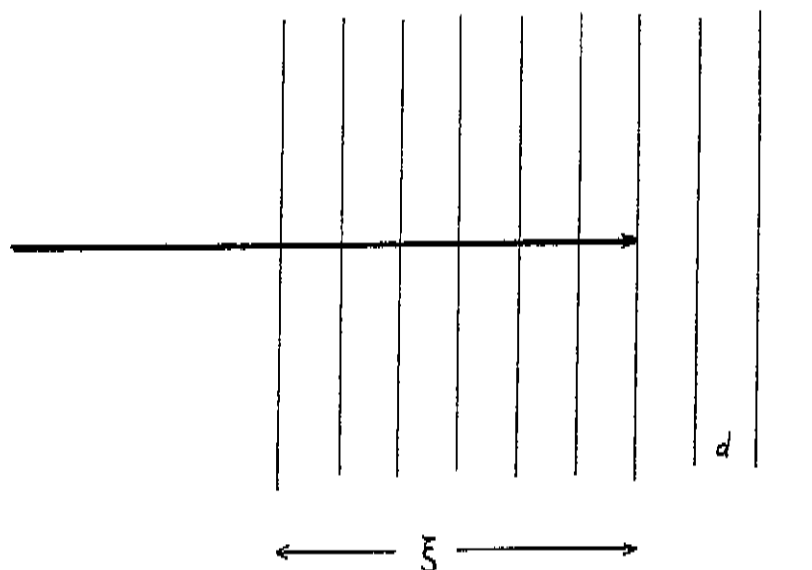
(16)

e $F(x)$ è esprimibile come integrale della densità di probabilità $f(x)$.



Esempio 3.

Un neutrone (o un fotone) passa attraverso una successione (infinita) di sottili fogli, a distanza d l'uno dall'altro, ognuno dei quali ha probabilità α ($0 < \alpha < 1$) di lasciar passare la particella. Possiamo misurare la distanza ξ tra il primo foglio e il foglio in cui avviene l'assorbimento.



ξ risulta una variabile casuale che può assumere i valori

- 0 - assorbimento al primo foglio
- d - assorbimento al secondo foglio
- $2d$ - assorbimento al terzo foglio.
- \vdots
- $(n-1)d$ - assorbimento al foglio n
- \vdots

Vogliamo determinare le corrispondenti probabilità e la probabilità cumulativa

$$F(x) = P\{\xi \leq x\}$$

Chiaramente, non essendoci alcun valore $\xi < 0$ avremo

$$F(x) = 0 \quad \text{per} \quad x < 0$$

Considerando ora il possibile assorbimento al primo foglio, $\xi = 0$, esso può avvenire con probabilità $1 - \alpha$, per cui

$$F(x) = (1 - \alpha) \quad \text{per} \quad 0 \leq x < d$$

Consideriamo la possibilità di un assorbimento al secondo foglio, $\xi = d$; perché avvenga un assorbimento al secondo foglio la particella deve avere superato indenne il primo, e questo può avvenire con probabilità α , pertanto la probabilità di assorbimento al secondo foglio risulta $\alpha(1 - \alpha)$. La probabilità cumulativa di essere assorbita in uno dei primi due fogli risulta quindi:

$$1 - \alpha + \alpha(1 - \alpha) = 1 - \alpha^2$$

e di conseguenza

$$F(x) = 1 - \alpha^2 \quad \text{per} \quad d \leq x < 2d$$

Passiamo ora all' n -esimo foglio. La probabilità di avere esattamente un assorbimento in tale foglio ($\xi = (n-1)d$) sarà il prodotto della corrispondente probabilità di assorbimento $1 - \alpha$ (abbiamo supposto tutti i fogli uguali) per la probabilità di attraversare indenne i primi $n-1$ fogli, cioè

$$\alpha^{n-1} (1 - \alpha)$$

La probabilità che il neutrone attraversi anche tale foglio risulterà invece α^n . In tal caso la particella sarà assorbita da uno dei fogli successivi: $\xi \geq nd$ e quindi:

$$P\{\xi \geq nd\} = \alpha^n$$

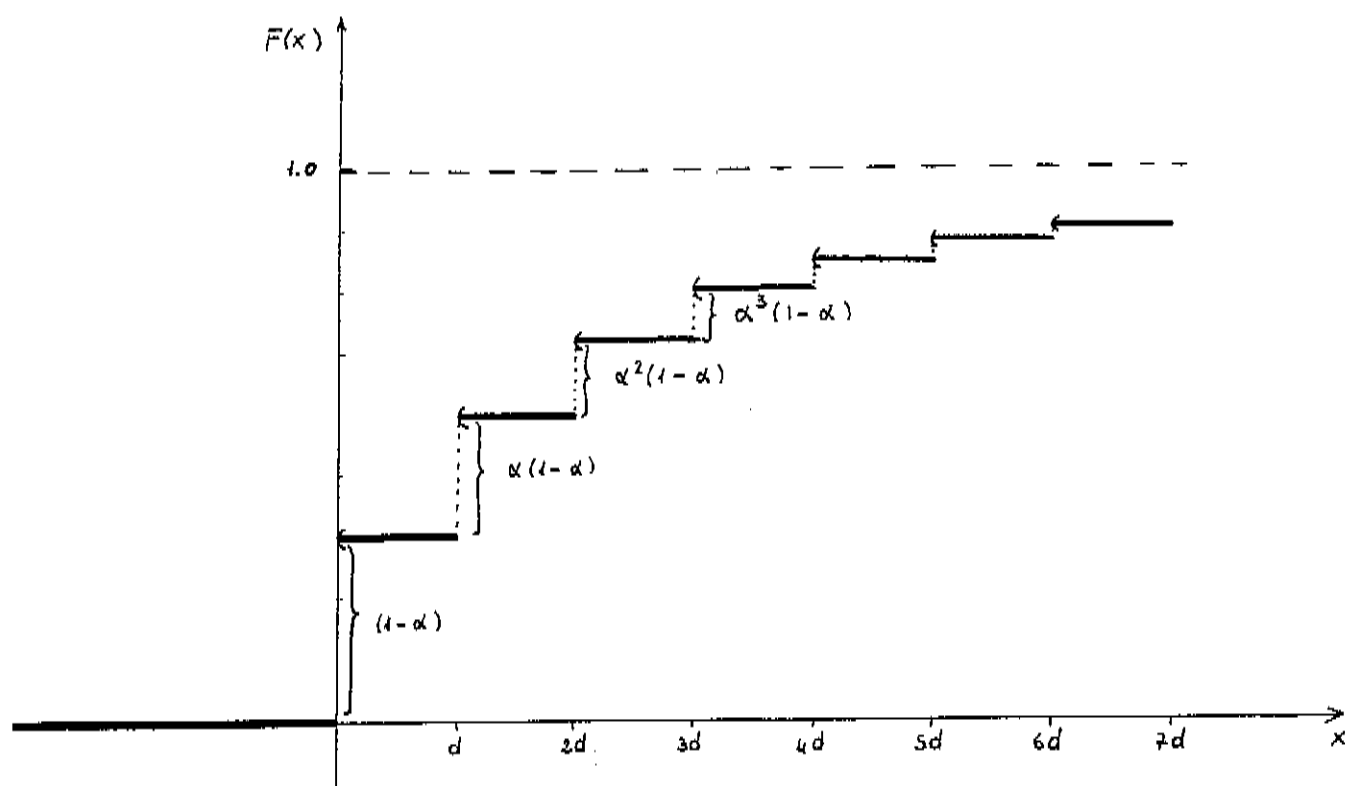
e possiamo dire

$$F(x) = 1 - \alpha^n \quad \text{per } (n-1)d \leq x < nd$$

cioè, in definitiva:

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - \alpha^n & \text{se } (n-1)d \leq x < nd \end{cases} \quad (2.14)$$

$n = 1, 2, \dots,$



Questa probabilità cumulativa $F(x)$ è una funzione a gradini con infiniti salti, e si noti la somiglianza con la probabilità cumulativa dell'esempio precedente. Sappati, notando che:

20

$$\alpha^n = e^{-n \ln \frac{1}{\alpha}} = e^{-\left(\frac{1}{d} \ln \frac{1}{\alpha}\right) nd}$$

possiamo introdurre un libero cammino medio dato da:

$$\lambda = \frac{d}{\ln \frac{1}{\alpha}} \quad (2.15)$$

(notiamo che, essendo $0 < \alpha < 1$, abbiamo $\ln \frac{1}{\alpha} > 0$).

Esempio 4.

dell'esempio precedente supponiamo che sia presente una differenza di potenziale elettrico in un circuito del tipo

$$\varphi(t) = \varphi_0 \sin \omega t \quad (2.16)$$

mentre la particella si muove attraverso i fogli. (il potenziale è sincronizzato con l'attraversamento del primo foglio).

Assumendo che il neutrone, prima dell'assorbimento, si muove con velocità costante, sia τ il tempo di volo del neutrone tra un foglio e l'altro, e sia $\theta = \omega\tau$. Allora il voltaggio φ al momento di assorbimento può avere il valore 0 con probabilità $1-\alpha$, il valore $\varphi_0 \sin \theta$ con probabilità $(1-\alpha)\alpha$, ..., il valore $\varphi_0 \sin n\theta$ con probabilità $\alpha^n(1-\alpha)$, ecc. In realtà $0, \theta, 2\theta, \dots, n\theta, \dots$ sono i possibili valori della fase al momento di assorbimento, con i corrispondenti valori di probabilità, $1-\alpha, (1-\alpha)\alpha, \dots, (1-\alpha)\alpha^2, \dots, (1-\alpha)\alpha^n, \dots$; e per fasi diverse, a causa della periodicità della funzione trigonometrica \sin , un medesimo voltaggio può presentarsi in piani diversi. In ogni caso possiamo esprimere la probabilità cumulativa

$$F(x) = P\{\varphi \leq x\}$$

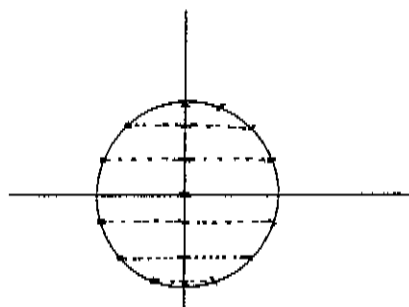
nella seguente maniera

$$F(x) = \begin{cases} 0 & x < -\varphi_0 \\ \sum_{\varphi_0 \sin(n\theta) \leq x} (1-\alpha)\alpha^n & -\varphi_0 \leq x < \varphi_0 \\ 1 & x \geq \varphi_0 \end{cases} \quad (2.17)$$

(22)

Se θ è una frazione razionale di π , ad esempio $\theta = \frac{\pi}{m}$, allora i valori possibili del potenziale sono un numero finito di valori (esattamente m)

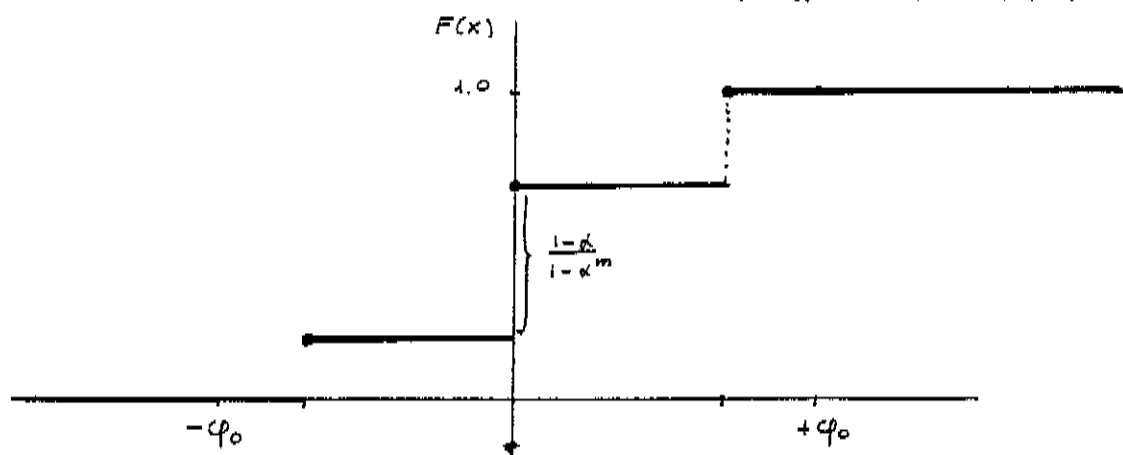
$$\varphi = \varphi_0 \sin n \frac{\pi}{m} \quad n = 0, 1, \dots, 2m-1$$



e si corrisponde ai valori $x = \varphi_0 \sin n \frac{\pi}{m}$ la probabilità cumulativa presenta un salto (finito) di ampiezza

$$(1 + \alpha^m) \sum_{k=0}^{\infty} (1 - \alpha) \alpha^{n+2km} = (1 - \alpha) \alpha^n \sum_{k=0}^{\infty} (\alpha^{2m})^k (1 + \alpha^m)$$

$$= (1 + \alpha^m) \frac{(1 - \alpha) \alpha^n}{1 - \alpha^{2m}} = \frac{\alpha^n + \alpha^{n+m}}{1 + \alpha + \dots + \alpha^{2m-1}}$$



Se θ non è un multiplo razionale di π allora $F(x)$ ha un numero infinito di salti, densamente distribuiti nell'intervallo $[-\varphi_0, \varphi_0]$, (la distanza tra un salto e l'altro si annulla, cioè i gradienti hanno ampiezza nulla, ma altezza finita.)

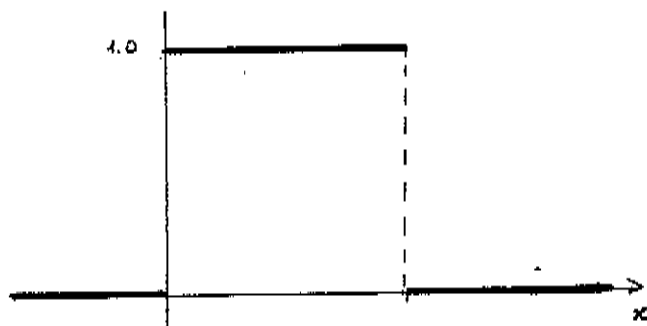
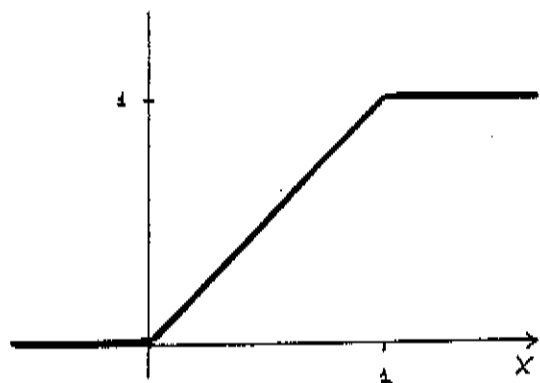
Esempio 5 La funzione di Cantor.

Supponiamo che un computer sia connesso con un dispositivo che generi, a domanda, un numero casuale ξ uniformemente distribuito nell'intervallo $[0, 1]$. Ciò significa che la variabile casuale ξ è collegata alla probabilità cumulativa:

$$F(x) = P\{\xi \leq x\} = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases} \quad (2.18)$$

oppure, equivalentemente, la probabilità associata alla variabile casuale ξ è data dalla densità di probabilità

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x < 1 \\ 0 & x \geq 1 \end{cases} \quad (2.19)$$



I valori della variabile casuale ξ sono del tipo

$$\xi = 0.a b c d \dots$$

con ogni cifra a, b, c, \dots 0 o 1 (in un computer le uniche cifre a disposizione sono i bit nulli o non nulli).

(84)

Assumiamo ora che il computer, dato il valore di ξ , lo trasformi in un numero η , ottenuto da ξ ripetendo due volte ogni cifra

$$\eta = 0.aabbccdd_{\dots}$$

η risulta anch'essa una variabile casuale che può assumere valori nell'intervallo $[0, 1]$. Possiamo valutare la probabilità cumulativa

$$F(y) = P\{\eta \leq y\}$$

Ogni numero η della forma sopra è sicuramente o minore di $0.01_2 = \frac{1}{4}$ (se $a=0$) oppure maggiore (o uguale) a 0.11_2 (se $a=1$) quindi

$$F(y) = \frac{1}{2} \quad \text{se} \quad \frac{1}{4} \leq y \leq \frac{3}{4}$$

(infatti tutto dipende dalla cifra binaria a che può assumere il valore 0 con probabilità $\frac{1}{2}$ oppure 1 con probabilità $\frac{1}{2}$, e i casi accettabili con y in tale intervallo sono quelli per cui $a=0$).

In maniera analoga, se y è compreso tra $0.0001_2 = \frac{1}{16}$ e $0.0011_2 = \frac{3}{16}$, $F(y)$ vale

$$F(y) = \frac{1}{4} = 0.01_2 \quad \frac{1}{16} \leq y \leq \frac{3}{16}$$

oppure, se y è compreso tra $0.1101_2 = \frac{13}{16}$ e $0.1111_2 = \frac{15}{16}$ abbiamo

$$F(y) = \frac{3}{4} = 0.11_2$$

In generale possiamo considerare che se y ha la forma

$$y = 0.aabbccdd\dots_2$$

(tutte le cifre ripetute un numero pari di volte) allora

$$\begin{aligned} F(y) &= \mathcal{P}\{ \eta \leq 0.aabbcc\dots \} = \\ &= \mathcal{P}\{ \xi \leq 0.abc\dots \} = .abc\dots \quad (2.20) \\ &\quad (y = .aabbcc\dots) \end{aligned}$$

Se invece y non ha tale forma, si troverà in un intervallo di costanza di F , e $F(y)$ assume uno dei valori sopra.

In fatti, se y ha la forma

$$y = 0.aabb\dots xx y_2\dots_2$$

con $y_2 \neq 2$ allora $y \in [0.aabb\dots xx 0_2, 0.aabb\dots xx 1_2]$
e

$$F(y) = 0.ab\dots x 1_2$$

Mostriamo che l'intervallo

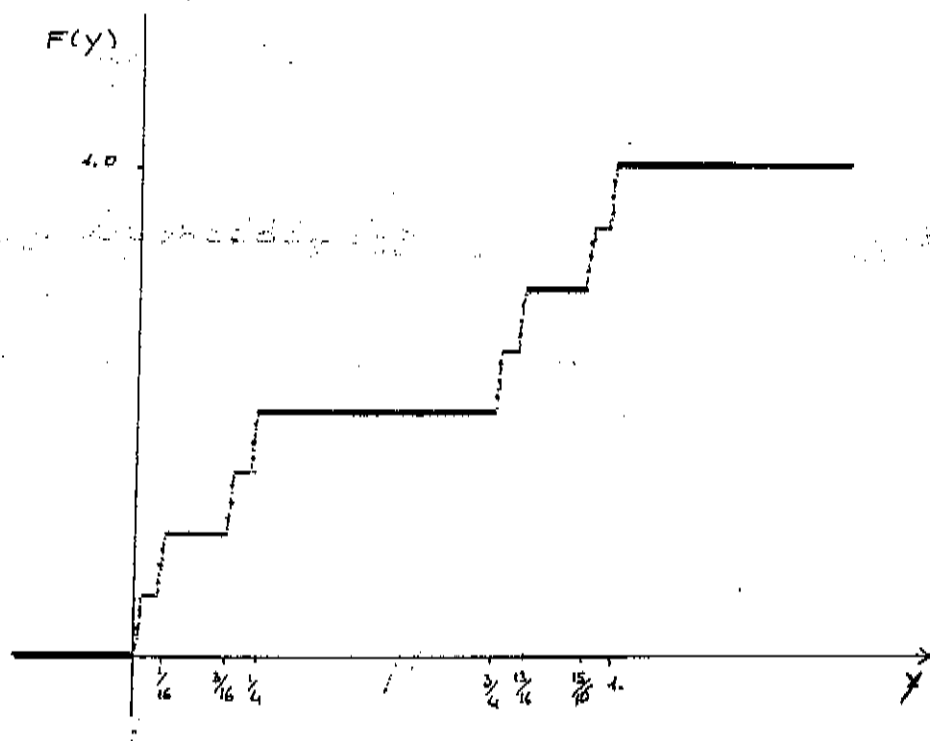
$$[0.a_0 a_0 \dots a_{n-1} a_{n-1} 0_2, 0.a_0 a_0 \dots a_{n-1} a_{n-1} 1_2]$$

ha ampiezza $0.\underbrace{00\dots 00}_{2n} 1_2 = \frac{1}{2^{2n+1}}$ e quindi la

somma di tutte le lunghezze degli intervalli lungo cui F è costante è:

$$\begin{aligned} \frac{1}{2} + 2 \cdot \frac{1}{8} + \dots + 2^n \frac{1}{2^{2n+1}} + \dots &= \\ &= \sum_{n=0}^{\infty} 2^n \frac{1}{2^{2n+1}} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = 1 \end{aligned}$$

e riempiamo tutto l'intervallo $[0, 1]$.



Questo fatto ci garantisce che la funzione $F(y)$ ha derivata nulla quasi dappertutto e per questo viene detta singolare. D'altra parte la funzione $F(y)$ è continua (sia da destra che da sinistra). Sappati se un punto y_0 ha la forma

$$y_0 = 0.a_0a_1\dots a_na_n\dots_2$$

allora

$$F(y_0) = 0.a_0\dots a_n\dots_2$$

e se $y \rightarrow y_0$, y passa attraverso punti di tale tipo e $F(y)$ si avvicina a $F(y_0)$. Se y_0 è interno a un intervallo di costanza di F , chiaramente questa risulta continua in y_0 .

Abbiamo a che fare con una funzione singolare continua. Essa costituisce un esempio di funzione di Cantor, continua, non costante, e con derivata nulla quasi ovunque.

Gli esempi precedenti mostrano sostanzialmente tutte le possibili caratteristiche di una probabilità cumulativa $F(x)$. Infatti, combinando due teoremi di decomposizione, uno dovuto a Jordan e uno dovuto a Lebesgue, si può dimostrare in generale che ogni funzione non decrescente $F(x)$ può essere scritta come somma di tre funzioni non decrescenti:

$$F(x) = F_1(x) + F_2(x) + F_3(x) \quad (2.21)$$

dove $F_1(x)$ è una funzione puramente a salti, con salti di ampiezza $p_1, p_2, \dots, p_n, \dots$ nei punti $x_1, x_2, \dots, x_n, \dots$ (non necessariamente ordinati), le ampiezze sono finite e $F_2(x)$ ha la forma

$$F_2(x) = \sum_{x_n \leq x} p_n \quad (\text{ogni } p_n > 0) \quad (2.22)$$

come negli esempi 1, 3 e 4.

$F_2(x)$ risulta invece una funzione continua; derivabile quasi ovunque, esprimibile come integrale (nel senso di Lebesgue) della sua derivata $f(x) \geq 0$:

$$F_2(x) = \int^x f(x) dx \quad (2.23)$$

$$f(x) = F_2'(x) \geq 0$$

Un esempio di funzione di tale tipo, che rientra nella classe delle funzioni assolutamente continue, è data dall'esempio 3.

$F_3(x)$ è invece una funzione singolar continua, cioè continua, non decrescente, con derivata quasi ovunque nulla. (chiaramente non può essere scritta come integrale della sua derivata). La funzione di Cantor dell'esempio 5 costituisce

un classico esempio di funzione di tale tipo.

Se una probabilità cumulativa è una funzione di puri salti, la corrispondente probabilità è detta atomica. Se i salti avvengono in punti isolati, come nell'esempio 1 e 3, essa viene detta discreta.

3. Medie e valori di aspettazione.

Supponiamo che $F(x)$ sia la probabilità cumulativa di una variabile random ξ e che desideriamo trovare il valore medio di ξ ottenuto da una lunga serie di misure. Consideriamo prima il caso in cui ξ è limitata ad un intervallo per cui $F(x) = 0$ per $x < a$ e $F(x) = 1$ per $x \geq b$, per cui i possibili valori di ξ sono confinati nell'intervallo $[a, b]$. Eseguiamo una partizione di $[a, b]$ in N intervalli delimitati dai punti x_i ($i = 0, \dots, N$) con

$$a = x_0 < x_1 < \dots < x_N = b$$

Sia ora x'_i un punto qualsiasi dell'intervallo $[x_{i-1}, x_i]$ ($i = 1, \dots, N$). La probabilità che ξ risulti in tale intervallo risulta $F(x_i) - F(x_{i-1})$, per cui una approssimazione per il valore medio di ξ risulta

$$\sum_{i=1}^N x'_i (F(x_i) - F(x_{i-1})) \quad (3.1)$$

e, nel limite di partizioni sempre più fini ($N \rightarrow \infty$), tale somma tende a (esiste il limite in questo tutto è limitato) all'integrale di Stieltjes

$$\int_a^b x dF(x) \quad (3.2)$$

L'integrale (3.2) viene allora detto la media (statistica) o valore di aspettazione di ξ e viene indicato con $E(\xi)$, $\langle \xi \rangle$ oppure μ . Nel caso generale in cui ξ non è a priori limitata ad un intervallo, il valore di aspettazione è dato da:

$$\mu = E(\xi) = \langle \xi \rangle = \int_{-\infty}^{+\infty} x dF(x) = \quad (3.3)$$

$$= \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b x dF(x)$$

a patto che tale limite esista. Se tale limite esiste ci si aspetta che il valore medio di una lunga serie di misure di una variabile ξ si stabilizzi attorno al valore $\langle \xi \rangle$ all'aumentare del numero di misure. In caso contrario il valor medio non si stabilizza all'aumentare del numero delle prove.

In maniera analoga possiamo definire il valore di aspettazione di una funzione $\varphi(\xi)$ della variabile casuale ξ (corrisponderebbe fisicamente ad una misura indiretta). Limitandoci al caso di funzioni continue $\varphi(x)$, possiamo definire

$$E(\varphi(\xi)) = \langle \varphi(\xi) \rangle = \int_{-\infty}^{+\infty} \varphi(x) dF(x) \quad (3.4)$$

ammesso che tale integrale (inteso sempre in senso generalizzato) esista.

Le prime applicazioni dei valori di aspettazione sono forse state nella teoria dei giochi di azzardo (legati ad eventi casuali). Infatti la vincita risulta in generale una funzione dell'evento, cioè una funzione $\varphi(\xi)$ della variabile casuale ξ , e generalmente minore è la probabilità dell'evento ξ , maggiore è la vincita $\varphi(\xi)$. In base alla definizione di valore di aspettazione si dà la definizione di gioco equo:

"Un gioco si dice equo quando la "posta", cioè l'ammontare di denaro necessario per partecipare

al gioco, è uguale al valore di aspettazione, detto anche speranza matematica, della vincita:

$$\text{posta} = \langle \text{vincita} \rangle$$

oppure, più in generale, quando il valore di aspettazione della vincita è uguale al valore di aspettazione delle perdite.

Se si esaminano i vari giochi, codificati e gestiti dalle case di gioco, dallo stato e così via, si noterà che la posta è sempre (o quasi) superiore alla speranza matematica della vincita, e in ogni caso non è mai inferiore.

Le compagnie di assicurazione utilizzano i concetti di speranza matematica, combinata con i rendimenti finanziari, per determinare gli importi dei premi assicurativi.

A proposito di giochi equi risulta famoso il "paradossso di Pietroburgo", ideato da Bernoulli all'Accademia di Pietroburgo. Supponiamo di lanciare una moneta da un centesimo con l'accordo che il lanciatore paghi all'avversario 1000 lire se viene testa. Se invece viene croce il lancio viene ripetuto e se viene testa viene pagata la somma di 2000 lire. Se viene ancora croce il lancio è ripetuto una terza volta e vengono pagate 4000 lire se esce testa. Cioè la vincita viene raddoppiata ogni volta che viene croce e pagata la prima volta che esce testa. Quanto dovrebbe mettere di posta l'avversario per avere il privilegio di giocare questo gioco unilaterale?

La risposta è che dovrebbero essere pagata qualsiasi somma, anche un miliardo, per ogni partita con la previsione di uscire comunque in vantaggio. Infatti, in ogni singola giocata vi è probabilità $\frac{1}{2}$ di

vincere 1000, $\frac{1}{4}$ di vincere 2000, $\frac{1}{8}$ di vincere 4000, e così via. La "speranza matematica" di vincita risulta quindi

$$1000 \cdot \frac{1}{2} + 2000 \cdot \frac{1}{4} + 4000 \cdot \frac{1}{8} + \dots = \dots$$

$$= 1000 \left(\frac{1}{2} + 2 \cdot \frac{1}{4} + \dots + 2^{n-1} \cdot \frac{1}{2^n} + \dots \right) = \infty$$

cioè una speranza di vincita infinita. Se paghiamo un milione per una partita abbiamo una forte probabilità di perdere, ma quanto è facile ricevere una vincita di poche migliaia di lire, ma questo fatto è bilanciato dalla possibilità di vincere una somma astronomica (anche se con bassissima probabilità) dopo una lunga serie di croci. Pertanto, avendo a disposizione (in teoria) un capitale illimitato, si può continuare a giocare con la previsione di vincere dopo un numero sufficientemente alto di partite.

Il paradosso di Pietroburgo si presenta in ogni sistema di gioco "a raddoppio", e la sua analisi completa conduce a un sacco di intricate deviazioni.*

Sostanzialmente possiamo riconoscere che l'origine del paradosso di Pietroburgo risiede nel fatto che la speranza matematica espressa dalla relazione (3.4) non è finita, cioè l'integrale, o serie che sia, non converge. Su questi casi, non ha senso, matematicamente, parlare di valori di aspettazione.

*) I giocatori del lotto che puntano sui numeri ritardatari raddoppiando ogni volta la puntata in caso di non uscita del numero ritardatario seguono una logica simile al paradosso di Pietroburgo, nell'erronea convinzione che un numero ritardatario abbia una probabilità maggiore di uscire.

Tra i valori di aspettazione rivestono particolare importanza le quantità $\langle \xi^k \rangle$ $k=1, 2, \dots$, quando questi esistono, e sono chiamati i momenti della distribuzione, spesso indicati con μ_k (momento di ordine k).

$$\mu_k = \langle \xi^k \rangle = \int_{-\infty}^{+\infty} x^k dF(x) \quad (3.5)$$

$k=1, 2, \dots$

Il primo momento ($k=1$) viene detto anche semplicemente media della distribuzione di probabilità:

$$\mu = \langle \xi \rangle = \int_{-\infty}^{+\infty} x dF(x)$$

mentre una importante combinazione dei primi due momenti è la varianza

$$\begin{aligned} \sigma^2 &= \langle (\xi - \mu)^2 \rangle = \langle \xi^2 \rangle - \langle \xi \rangle^2 \\ &= \mu_2 - \mu_1^2 \end{aligned} \quad (3.6)$$

σ stesso viene detta deviazione standard della variabile casuale ξ .

Possiamo dare significato anche al caso $k=0$ definendo

$$\mu_0 = 1 = \int_{-\infty}^{+\infty} dF(x) \quad (3.7)$$

nel caso di variabili casuali ξ limitate (che possono assumere valori solo all'interno di un intervallo limitato $[a, b]$) i momenti μ_k esistono e sono finiti a tutti gli ordini.

I momenti μ_k costituiscono una sorta di firma della distribuzione di probabilità. Un problema classico importante, il problema dei momenti, è appunto quello di determinare

(34)

la probabilità cumulativa F (cioè la distribuzione di probabilità) quando sono noti tutti i momenti μ_k .

Un altro valore di aspettazione importante, che gioca un ruolo importante nel teorema del limite centrale, è dato dalla funzione complessa

$$\chi(\lambda) = \langle e^{-i\lambda\xi} \rangle = \int_{-\infty}^{+\infty} e^{-i\lambda x} dF(x) \quad (3.8)$$

detta funzione caratteristica della distribuzione di probabilità. Nel caso in cui $F(x)$ sia una funzione assolutamente continua, con derivata $f(x) = F'(x) \geq 0$ ben definita e per cui

$$F(x) = \int_{-\infty}^x f(x) dx$$

la funzione caratteristica risulta, a parte fattori di normalizzazione $\sqrt{2\pi}$, la trasformata di Fourier della densità di probabilità $f(x)$.

Formalmente abbiamo un legame tra la funzione caratteristica e i momenti μ_k . Infatti, quando sono leciti matematicamente tutti i passaggi:

$$\begin{aligned} \chi(\lambda) &= \int_{-\infty}^{+\infty} e^{-i\lambda x} dF(x) = \\ &= \int_{-\infty}^{+\infty} \sum_{k=0}^{\infty} \frac{1}{k!} (-i\lambda x)^k dF(x) = \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} (-i\lambda)^k \int_{-\infty}^{+\infty} x^k dF(x) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \mu_k (-i\lambda)^k \end{aligned} \quad (3.9)$$

Esempio 6.

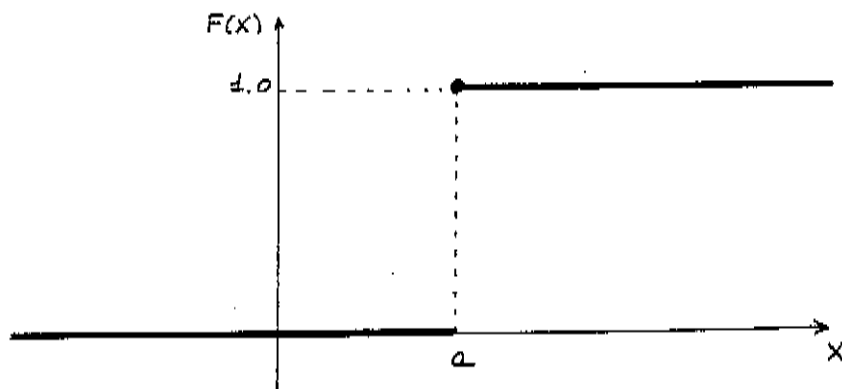
A titolo di esempio consideriamo un caso a prima vista banale, ma importante da un punto di vista concettuale. Supponiamo che il risultato di una misura sia sempre lo stesso numero, cioè abbiamo una variabile casuale che ha un solo valore $\xi = a$ ammesso. Abbiamo a che fare, da un punto di vista probabilistico con un evento certo, esattamente determinato, la probabilità di avere $\xi = a$ è esattamente 1, e la probabilità di avere un qualsiasi altro valore è nulla

$$P \{ \xi = a \} = 1 \quad (3.10)$$

$$P \{ \xi \neq a \} = 0$$

In tal caso la probabilità cumulativa è un semplice scalino di ampiezza 1:

$$F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases} \quad (3.11)$$



Vista in termini di densità di probabilità questa funzione (detta $\theta(x-a)$ o teta di Heavyside) definisce la distribuzione delta di Dirac:

$$\delta(x-a) = F'(x) = \frac{d}{dx} \theta(x-a) \quad (3.12)$$

una volta risolte le ambiguità nella definizione di derivata (non è certamente una derivata ordinaria).

Vogliamo determinare la "firma" di una distribuzione di probabilità di tale tipo. In base alle definizioni poste abbiamo

$$\begin{aligned}\langle \varphi(\xi) \rangle &= \int_{-\infty}^{+\infty} \varphi(x) dF(x) \\ &= \varphi(a)\end{aligned}\tag{3.13}$$

per cui, in particolare

$$\mu_k = \langle \xi^k \rangle = a^k\tag{3.14}$$

i momenti sono esattamente delle potenze del numero, a (l'evento certo), e la funzione caratteristica risulta

$$\chi(\lambda) = e^{-i\lambda a}\tag{3.15}$$

un'onda piana in termini di λ .

In particolare possiamo notare che la media μ coincide con a , mentre la varianza σ^2 risulta nulla

$$\mu = a$$

$$\sigma^2 = 0$$

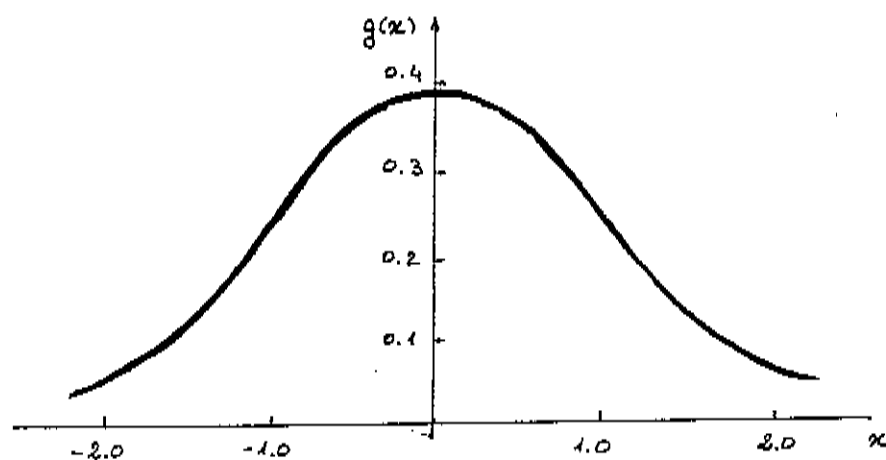
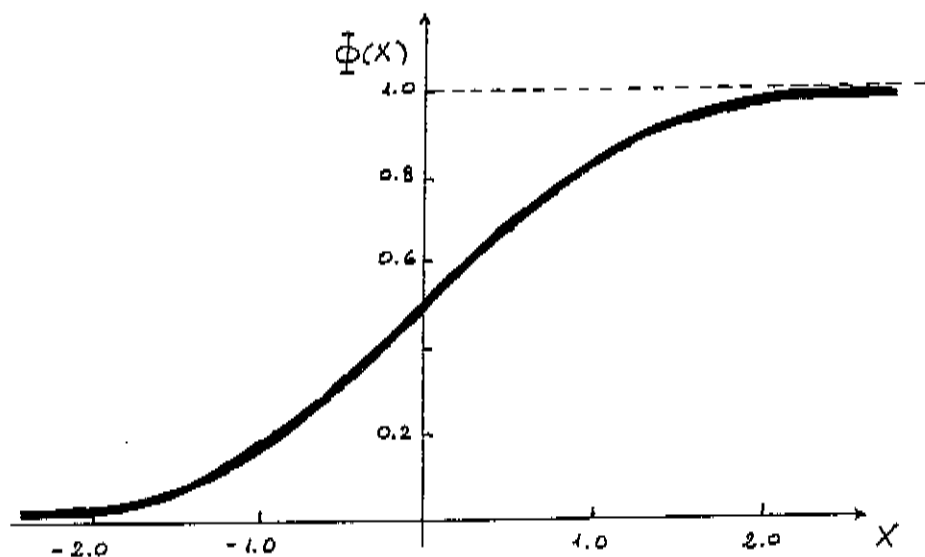
4. La distribuzione normale o di Gauss.

La distribuzione normale, o di Gauss, o Gaussiana, è definita come la distribuzione di probabilità descritta dalla seguente probabilità cumulativa

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx \quad (4.1)$$

o, equivalentemente, dalla densità di probabilità di Gauss:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (4.2)$$



(36)

Possiamo vedere la "firma" della distribuzione gaussiana.
Abbiamo ovviamente:

$$\mu_0 = \phi(\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1 \quad (4.3)$$

mentre i momenti μ_k di ordine dispari sono chiaramente nulli per antisimmetria della funzione integranda.

$$\mu_{2k+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k+1} e^{-\frac{x^2}{2}} dx = 0 \quad (4.4)$$

$k = 0, 1, 2, \dots$

Per valutare i momenti di ordine pari consideriamo l'integrale

$$\begin{aligned} V(\alpha) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\alpha \frac{x^2}{2}} dx = \quad (\alpha > 0) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(\sqrt{\alpha} x)^2}{2}} \frac{d(\sqrt{\alpha} x)}{\sqrt{\alpha}} \\ &= \frac{1}{\sqrt{\alpha}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\sqrt{\alpha}} = \alpha^{-1/2} \end{aligned}$$

Pertanto:

$$\begin{aligned} \frac{d^k}{d\alpha^k} V(\alpha) &= \left(-\frac{1}{2}\right) \left(-\frac{3}{2}\right) \dots \left(-\frac{1}{2} - k + 1\right) \alpha^{-1/2 - k} \\ &= \frac{(-1)^k}{2^k} 1 \cdot 3 \cdot \dots (2k-1) \alpha^{-1/2 - k} \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(-\frac{x^2}{2}\right)^k e^{-\alpha \frac{x^2}{2}} dx \\ &= \frac{(-1)^k}{2^k} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k} e^{-\alpha \frac{x^2}{2}} dx \end{aligned}$$

cioè:

$$\underbrace{1 \cdot 3 \cdot 5 \cdots (2k-1)}_{k \text{ fattori}} \alpha^{-\frac{1}{2}-k} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k} e^{-\alpha \frac{x^2}{2}} dx$$

k fattori

e, ponendo ora $\alpha = 1$,

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2k} e^{-\frac{x^2}{2}} dx &= 1 \cdot 3 \cdot 5 \cdots (2k-1) \\ &= (2k-1)!! \end{aligned}$$

peranto

$$\left\{ \begin{array}{l} \mu_0 = 1 \\ \mu_{2k-1} = 0 \\ \mu_{2k} = (2k-1)!! \end{array} \right. \quad (4.5)$$

Mediante le relazioni (4.5) è possibile valutare anche la funzione caratteristica della distribuzione di Gauss:

$$\begin{aligned} \chi(\lambda) &= \sum_{k=0}^{\infty} \frac{1}{k!} \mu_k (-i\lambda)^k = \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \mu_{2k} (-i\lambda)^{2k} = \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \mu_{2k} (-\lambda^2)^k \\ &= 1 + \sum_{k=1}^{\infty} \frac{1}{(2k)!} \frac{1}{(2k-1)!!} (-\lambda^2)^k (2k-1)!! \\ &= 1 + \sum_{k=1}^{\infty} \frac{1}{(2k)(2k-2)\cdots(2)} (-\lambda^2)^k \end{aligned}$$

$$\begin{aligned}
 &= 1 + \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot \frac{1}{k(k-1)\dots 1} (-\lambda^2)^k \\
 &= \sum_{k=0}^{\infty} \frac{1}{k!} \left(-\frac{\lambda^2}{2}\right)^k \\
 &= e^{-\frac{\lambda^2}{2}}
 \end{aligned}$$

cioè

$$X(\lambda) = e^{-\frac{\lambda^2}{2}} \quad (4.6)$$

(Sostanzialmente la funzione caratteristica coincide con la densità di probabilità gaussiana stessa).

La distribuzione gaussiana (4.1) è nota anche come curva degli errori o funzione degli errori $\text{Erf}(X)$ per la funzione che essa ricopre nella teoria degli errori di misurazione. La funzione $\phi(X)$ non ha una espressione in termini finiti e la si può trovare valutata con estrema precisione su molte tabelle. Quasi tutti i manuali con tabelle utili da un punto di vista scientifico riportano tabulata la funzione $\phi(X)$ al variare di X . In questo modo è possibile stimare ad esempio che:

$$\phi(-0.67449) = \frac{1}{4}$$

$$\phi(+0.67449) = \frac{3}{4}$$

da cui

$$P\{-0.67449 \leq \xi \leq 0.67449\} = \frac{1}{2}$$

La distribuzione (4.2) è detta normalizzata in quanto è a media nulla (centrata nel punto $x=0$) e varianza unitaria. In generale possiamo dire che una variabile casuale ξ è distribuita secondo la distribuzione normale, o di Gauss, con media μ e varianza σ^2 se (*)

$$P\{\xi \leq x\} = G(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad (4.7)$$

cioè:

$$\begin{aligned} G(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned} \quad (t = \frac{x-\mu}{\sigma})$$

con una densità di probabilità

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4.8)$$

detta ancora densità di probabilità gaussiana o normale.

Questa densità di probabilità è definita da soli due parametri μ e σ^2 che assumono proprio il significato di media e varianza (in termini probabilistici o statistici) della distribuzione di probabilità.

$$\mu = \langle \xi \rangle = \int_{-\infty}^{+\infty} x g(x) dx = \mu_1 \quad (4.9)$$

$$\sigma^2 = \langle \xi^2 \rangle - \langle \xi \rangle^2 = \mu_2 - \mu_1^2 \quad (4.10)$$

(*) Notiamo che nella (4.7) consideriamo $\sigma > 0$, cioè

$$\sigma = +\sqrt{\sigma^2}$$

(40)

La "firma" della distribuzione generale di Gauss non è più così semplice in generale. Risulta però relativamente semplice valutare la funzione caratteristica

$$\begin{aligned}
 X(\lambda) &= \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{+\infty} e^{-i\lambda x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-i\lambda(\mu+\sigma t)} e^{-\frac{t^2}{2}} dt \\
 &= \frac{1}{\sqrt{2\pi}} e^{-i\lambda\mu} \int_{-\infty}^{+\infty} e^{-i(\sigma\lambda)t - \frac{t^2}{2}} dt
 \end{aligned}$$

e quindi, dalla relazione (4.6)

$$\begin{aligned}
 X(\lambda) &= e^{-i\lambda\mu - \frac{\sigma^2\lambda^2}{2}} \\
 &= e^{-\frac{\mu^2}{2\sigma^2}} e^{-\frac{\sigma^2}{2}(\lambda + i\frac{\mu}{\sigma^2})^2}
 \end{aligned} \tag{4.11}$$

I momenti possono essere valutati dalla (4.11) usando la relazione: (deducibile dallo sviluppo in serie (3.8))

$$\mu_k = \left(i \frac{d}{d\lambda} \right)^k X(\lambda) \Big|_{\lambda=0} \tag{4.12}$$

Abbiamo

$$\begin{aligned}
 \mu_k &= \left\{ \left[i \frac{d}{d\lambda} \right]^k e^{-i\lambda\mu - \frac{\sigma^2\lambda^2}{2}} \right\} \Big|_{\lambda=0} \\
 &= e^{-\frac{\mu^2}{2\sigma^2}} \left\{ \left[i \frac{d}{d\lambda} \right]^k e^{-\frac{1}{2}(\sigma\lambda + i\frac{\mu}{\sigma})^2} \right\} \Big|_{\lambda=0} \\
 &= (i\sigma)^k e^{-\frac{\mu^2}{2\sigma^2}} \left\{ \left[\frac{d}{d(\sigma\lambda + i\frac{\mu}{\sigma})} \right]^k e^{-\frac{1}{2}(\sigma\lambda + i\frac{\mu}{\sigma})^2} \right\} \Big|_{\lambda=0}
 \end{aligned}$$

$$= (i\sigma)^k e^{-\frac{\mu^2}{2\sigma^2}} \left\{ \frac{d^k}{du^k} e^{-\frac{u^2}{2}} \right\} / u = i \frac{\mu}{\sigma}$$

$$= (i\sigma)^k \left\{ e^{\frac{u^2}{2}} \frac{d^k}{du^k} e^{-\frac{u^2}{2}} \right\} / u = i \frac{\mu}{\sigma}$$

Ricordando la definizione di polinomio di Hermite di grado n

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \quad (4.13)$$

abbiamo

$$\mu_k = (-1)^k \left(-i \frac{\sigma}{\sqrt{2}} \right)^k \left\{ e^{u^2} \frac{d^k}{du^k} e^{-u^2} \right\} / u = i \frac{\mu}{\sqrt{2}\sigma}$$

cioè:

$$\mu_k = \left(-i \frac{\sigma}{\sqrt{2}} \right)^k H_k \left(i \frac{\mu}{\sqrt{2}\sigma} \right) \quad (4.14)$$

Notiamo che i polinomi di Hermite hanno parità ben definita

$$H_n(-x) = (-1)^n H_n(x) \quad (4.15)$$

per cui i momenti μ_k risultano reali e polinomi omogenei in μ e σ . Infatti

$$\begin{aligned} \text{Im } \mu_k &= \frac{1}{2i} (\mu_k - \bar{\mu}_k) = \\ &= \frac{1}{2i} \left[\left(-i \frac{\sigma}{\sqrt{2}} \right)^k H_k \left(i \frac{\mu}{\sqrt{2}\sigma} \right) - \left(+i \frac{\sigma}{\sqrt{2}} \right)^k H_k \left(-i \frac{\mu}{\sqrt{2}\sigma} \right) \right] \\ &= \frac{1}{2i} \left[\left(-i \frac{\sigma}{\sqrt{2}} \right)^k H_k \left(\frac{\mu}{-i\sigma\sqrt{2}} \right) - \left(-i \frac{\sigma}{\sqrt{2}} \right)^k H_k \left(\frac{\mu}{-i\sigma\sqrt{2}} \right) \right] \\ &= 0 \end{aligned}$$

Consultando una qualsiasi tavola di polinomi di Hermite abbiamo ad esempio

$$H_0(x) = 1$$

$$H_1(x) = 2x$$

$$H_2(x) = 4x^2 - 2$$

$$H_3(x) = 8x^3 - 12x$$

$$H_4(x) = 16x^4 - 48x^2 + 12$$

da cui:

$$\mu_0 = 1$$

$$\mu_1 = \mu$$

$$\mu_2 = \mu^2 + \sigma^2$$

$$\mu_3 = \mu^3 + 3\mu\sigma^2$$

$$\mu_4 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

(Notiamo che, poiché il polinomio di Hermite ha come termine di grado più elevato $(2x)^n$, μ_n ha come termine di grado più elevato in μ il semplice termine μ^n , con coefficiente 1).

Graficamente la distribuzione di Gauss con media μ e varianza σ^2 è simile alle figure già presentate. L'unica differenza è che la distribuzione è centrata attorno al punto $x = \mu$ e scalata di un fattore di scala σ lungo l'asse x (per la densità di probabilità abbiamo un analogo fattore di scala σ^{-1} lungo l'asse y , per normalizzare a 1 l'integrale).

Dalla proprietà dei polinomi di Hermite:

$$H_n(x) = (2x)^n - \frac{n(n-1)}{1!} (2x)^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2!} (2x)^{n-4} - \dots + \frac{(-1)^k}{k!} n(n-1)\dots(n-2k+1) (2x)^{n-2k} + \dots \quad (4.16)$$

abbiamo:

$$\begin{aligned} \left(-\frac{i\sigma}{\sqrt{2}}\right)^n \left(2i\frac{\mu}{\sigma\sqrt{2}}\right)^{n-2k} &= (-1)^n i^{2n} (-i)^{2k} \left(\frac{1}{2}\right)^k \mu^{n-2k} \sigma^{2k} \\ &= (-1)^k \left(\frac{1}{2}\right)^k \mu^{n-2k} \sigma^{2k} \end{aligned}$$

$$\begin{aligned} \mu_n &= \mu^n + \frac{n(n-1)}{1! 2} \mu^{n-2} \sigma^2 + \frac{n(n-1)(n-2)(n-3)}{2! 4} \mu^{n-4} \sigma^4 \\ &\quad + \dots + \frac{n(n-1)\dots(n-2k+1)}{k! 2^k} \mu^{n-2k} \sigma^{2k} + \dots \end{aligned} \quad (4.17)$$

cioè

$$\mu_n = \sum_{2k \leq n} \frac{n(n-1)\dots(n-2k+1)}{k! 2^k} \mu^{n-2k} \sigma^{2k} \quad (4.18)$$

La relazione (4.18) costituisce la "firma", assieme alla (4.11) della distribuzione normale di Gauss.

Osserviamo che, per $\sigma \rightarrow 0$ la densità di probabilità risulta sempre più "piccata" attorno al valore $x = \mu$ e la probabilità cumulativa $G(x)$ tende ad uno "scalino" di altezza unitaria attorno al punto $x = \mu$

$$\lim_{\sigma \rightarrow 0^+} \Phi\left(\frac{x-\mu}{\sigma}\right) = \begin{cases} \Phi(+\infty) = 1 & x > \mu \\ \Phi(-\infty) = 0 & x < \mu \end{cases} \quad (4.19)$$

mentre per $x = \mu$ abbiamo

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(0) = \frac{1}{2} \quad (4.20)$$

(44)

anche per $\sigma \rightarrow 0^+$.

In questo caso vediamo che in μ_n sopravvive solo il termine μ^n ottenendo la firma dell'evento certo $\xi = \mu$.

$$\lim_{\sigma \rightarrow 0^+} \mu_n = \mu^n$$

(4.21)

$$\lim_{\sigma \rightarrow 0^+} X(\lambda) = e^{-i\lambda\mu}$$

mentre, per $\mu = 0$ otteniamo una firma più semplice, simile al caso normalizzato $\phi(x)$:

$$\begin{aligned} \lim_{\mu \rightarrow 0} X(\lambda) &= e^{-\frac{1}{2}\sigma^2\lambda^2} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \sigma^{2k} \lambda^{2k} \frac{1}{2^k} \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \frac{\sigma^{2k}}{2^k} (-i\lambda)^{2k} \\ &= \sum_{k=0}^{\infty} \frac{1}{(2k)!} \frac{(2k)!}{k! 2^k} \sigma^{2k} (-i\lambda)^{2k} \end{aligned}$$

da cui, ($n=0,1,\dots$)

$$\mu_{2n+1} = 0$$

$$\mu_{2n} = \frac{(2n)!}{n! 2^n} \sigma^{2n} = \underbrace{(2n-1)(2n-3)\dots 3 \cdot 1}_{n \text{ fattori}} \sigma^{2n}$$

(4.22)

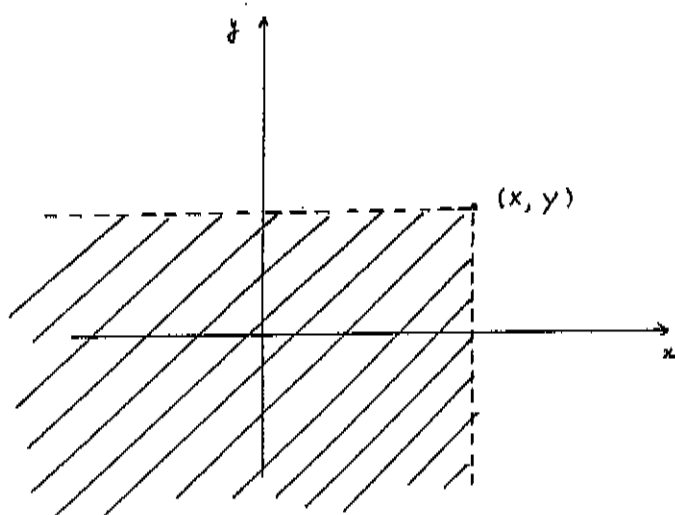
con la convenzione $\mu_0 = 1$.

5. Distribuzioni bivariate e multivariate.

Se il risultato di un esperimento è espresso da due numeri x, y , questi sono i valori assunti da una coppia di variabili casuali ξ, η e la corrispondente probabilità è detta bivariata.
In questo caso la probabilità cumulativa congiunta è data da

$$F(x, y) = P\{\xi \leq x \text{ e } \eta \leq y\} \quad (5.1)$$

ed esprime sostanzialmente la frazione di esperimenti il cui risultato, un punto nel piano, giace nel quadrante inferiore sinistro rispetto al punto (x, y)



Una qualsiasi probabilità connessa a tali eventi può essere espressa tramite la funzione F . In particolare avremo

$$\begin{aligned} P\{x_1 < \xi \leq x_2 \text{ e } y_1 < \eta \leq y_2\} &= \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \end{aligned} \quad (5.2)$$

Se con \square indichiamo la regione rettangolare:

$$\square = \{(x, y); x_1 < x \leq x_2 \text{ e } y_1 < y \leq y_2\}$$

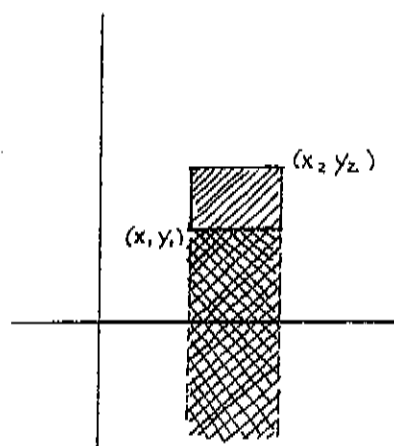
N.B. Assumiamo $x_1 < x_2$, $y_1 < y_2$

(46)

allora possiamo riscrivere la (5.2) nella forma più compatta

$$\mathcal{P}\{(\xi, \eta) \in \square\} = F(\square) \quad (5.3)$$

dove $F(\square)$ è definito dal membro destro della (5.2)



$$F(x_2, y_2) - F(x_1, y_2) = \mathcal{P}\{x_1 < \xi \leq x_2, \eta \leq y_2\}$$

$$F(x_2, y_1) - F(x_1, y_1) = \mathcal{P}\{x_1 < \xi \leq x_2, \eta \leq y_1\}$$

$$F(\square) = \mathcal{P}\{\square\}$$

Il fatto che $F(x, y)$ sia interpretabile come probabilità ha delle conseguenze ovvie. Chiaramente si deve avere $F(x, y) \geq 0$ e

$$F(-\infty, -\infty) = 0 \quad F(+\infty, +\infty) = 1 \quad (5.4)$$

ma anche:

$$\begin{aligned} F(x, -\infty) &= 0 & \forall x \\ F(-\infty, y) &= 0 & \forall y \end{aligned} \quad (5.5)$$

mentre la richiesta di monotonia del caso unidimensionale viene rimpiazzata dalla richiesta

$$F(\square) \geq 0 \quad \forall x_1 \leq x_2, y_1 \leq y_2 \quad (5.6)$$

e con tale richiesta si includono anche i casi limite

$$F(x_2, y_2) - F(x_1, y_2) \geq 0 \quad x_1 \leq x_2, \forall y_2$$

$$F(x_2, y_2) - F(x_2, y_1) \geq 0 \quad y_1 \leq y_2, \quad \forall x_2$$

ottenuti rispettivamente dalla (5.6) e (5.5) per $y_1 \rightarrow -\infty$ e $x_1 \rightarrow -\infty$. Su queste condizioni $F(x, y)$ è detta una funzione non decrescente delle due variabili (x, y) .

La definizione di $F(\square)$ permette di definire gli integrali e i valori medi. Considerando per semplicità una regione rettangolare $[a, b] \times [c, d]$ del piano xy , prendiamo una suddivisione di tale regione in rettangoli \square_j e consideriamo la somma

$$\sum_j \varphi(x_j, y_j) F(\square_j) \quad (x_j, y_j) \in \square_j$$

con φ funzione continua. Tale somma, nel limite di una suddivisione sempre più fine, converge all'integrale

$$\int_a^b \int_c^d \varphi(x, y) d^2 F(x, y)$$

(Questa integrazione è detta di Riemann-Stieltjes). Passando in maniera ovvia a tutto il piano possiamo definire i valori di aspettazione o valori medi:

$$\langle \varphi(\xi, \eta) \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) d^2 F(x, y) \quad (5.7)$$

del caso in cui esista

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

e sia continua (a tratti) e non nulla quasi ovunque, detta densità di probabilità, allora F è detta assolutamente continua e vale:

$$f(x, y) \geq 0$$

$$F(x, y) = \iint_{\substack{x \leq X \\ y \leq Y}} f(x, y) dx dy \quad (5.8)$$

$$\langle \varphi(\xi, \eta) \rangle = \int_{\mathbb{R}^2} \varphi(x, y) f(x, y) dx dy \quad (5.9)$$

I valori medi

$$\langle \xi^k \eta^l \rangle = \int_{\mathbb{R}^2} x^k y^l d^2 F(x, y) \quad (5.10)$$

sono detti i momenti della distribuzione bivariata, quando tali integrali sono convergenti. Mediante i momenti di ordine 2 (e uno) si costruisce la matrice di covarianza, con componenti

$$\begin{aligned} g_{11} &= \langle (\xi - \langle \xi \rangle)^2 \rangle = \langle \xi^2 \rangle - \langle \xi \rangle^2 \\ g_{22} &= \langle (\eta - \langle \eta \rangle)^2 \rangle = \langle \eta^2 \rangle - \langle \eta \rangle^2 \\ g_{12} &= g_{21} = \langle (\xi - \langle \xi \rangle)(\eta - \langle \eta \rangle) \rangle \\ &= \langle \xi \eta \rangle - \langle \xi \rangle \langle \eta \rangle \end{aligned} \quad (5.11)$$

e il coefficiente di correlazione

$$\rho = \frac{g_{12}}{\sqrt{g_{11} g_{22}}} \quad (5.12)$$

(definito quando $g_{11} g_{22} \neq 0$). In accordo con la disuguaglianza di Schwartz ρ verifica la condizione

$$-1 \leq \rho \leq +1 \quad (5.13)$$

Se $\rho = \pm 1$ ξ e η sono completamente correlati; e i punti x, y che possono risultare da un esperimento giacciono su una linea retta attraverso il punto di coordinate $(\langle \xi \rangle, \langle \eta \rangle)$.

Se ξ e η sono variabili indipendenti la funzione di probabilità congiunta ha la forma

$$F(x, y) = F_1(x) F_2(y) \quad (\text{indipendenza}) \quad (5.14)$$

oppure, nel caso questa derivi da una densità di probabilità:

$$f(x, y) = f_1(x) f_2(y) \quad (5.15)$$

In questo caso il coefficiente di correlazione ρ risulta nullo $\rho = 0$ (assenza di correlazione). Comunque notiamo che ρ può essere nullo anche quando ξ e η non sono indipendenti; un esempio è dato dalla distribuzione con densità di probabilità

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{per } x^2 + y^2 < 1 \\ 0 & \text{per } x^2 + y^2 > 1 \end{cases}$$

dove f_{12} è nullo per ragioni di simmetria, ma $f(x, y)$ non è fattorizzabile nel prodotto di due funzioni $f_1(x) f_2(y)$.

Analogamente ai momenti possiamo definire la funzione caratteristica di una distribuzione $\lambda = (\lambda_1, \lambda_2) \quad x = (x_1, x_2)$

$$X(\lambda) = \iint e^{-i \lambda \cdot x} d^2 F(x_1, x_2) \quad (5.16)$$

che, nel caso di probabilità assolutamente continue, esprime la

trasformata di Fourier della densità di probabilità $f(x_1, x_2)$

Tutto quanto detto sopra può essere generalizzato a n dimensioni in maniera banale. Se abbiamo una n -upla

$\xi = (\xi_1, \dots, \xi_n)$ di variabili casuali, la probabilità cumulativa (congiunta)

$$F(x) = F(x_1, \dots, x_n)$$

$$F(x_1, \dots, x_n) = P \{ \xi_1 \leq x_1, \dots, \xi_n \leq x_n \} \quad (5.17)$$

Giudichiamo con

$$\square = \{ x : a_1 < x_1 \leq b_1, \dots, a_n < x_n \leq b_n \} \quad (5.18)$$

e definiamo

$$F(\square) = P \{ \xi \in \square \} \quad (5.19)$$

la cui formula esplicita è una generalizzazione della (5.2).

Detto v un vertice di \square , come un punto le cui coordinate sono uguali ad a_j oppure a b_j , e detto $N_+(v)$ il numero di a (estremi inferiori) tra le coordinate di v , possiamo scrivere

$$F(\square) = \sum_v (-1)^{N_+(v)} F(v) \quad (5.20)$$

Allora dalla interpretazione probabilistica di F risulta ovvio che

- $F(x)$ è non decrescente, cioè:

$$F(\square) \geq 0 \quad \text{per ogni } \square$$

- $F(x)$ è normalizzata:

$$F(\infty, \infty, \dots, \infty) = 1$$

$$F(x_1, \dots, x_n) = 0 \quad \text{se una coordinata } x_j = -\infty$$

Tramite la (5.20) possiamo allora definire gli integrali

$$\int_{\square} \varphi(x) d^n F(x)$$

e i valori medi, quando gli integrali esistono

$$\langle \varphi(\xi) \rangle = \int_{\mathbb{R}^n} \varphi(x) d^n F(x) \quad (5.21)$$

e in particolare possiamo definire i momenti

$$\langle \xi^k \rangle = \langle \xi_1^{k_1} \dots \xi_n^{k_n} \rangle = \int_{\mathbb{R}^n} x_1^{k_1} \dots x_n^{k_n} d^n F(x_1, \dots, x_n) \quad (5.22)$$

e la funzione caratteristica

$$X(\lambda) = \int_{\mathbb{R}^n} e^{-i \lambda \cdot x} d^n F(x) \quad (5.23)$$

Tramite più variabili casuali ξ_1, \dots, ξ_n possiamo costruire nuove variabili casuali come funzioni di queste

$$\eta = g(\xi_1, \dots, \xi_n) \quad (5.24)$$

e dalla distribuzione di probabilità per le n variabili casuali ξ_1, \dots, ξ_n possiamo dedurre la distribuzione di probabilità per la nuova variabile casuale η :

$$\begin{aligned}
 G(y) &= \mathbb{P} \{ \eta \leq y \} = \\
 &= \mathbb{P} \{ g(\xi_1, \dots, \xi_n) \leq y \} = \\
 &= \iint \dots \int_{\{g(x) \leq y\}} d^n F(x)
 \end{aligned}
 \tag{5.25}$$

e, se possiamo parlare di densità di probabilità

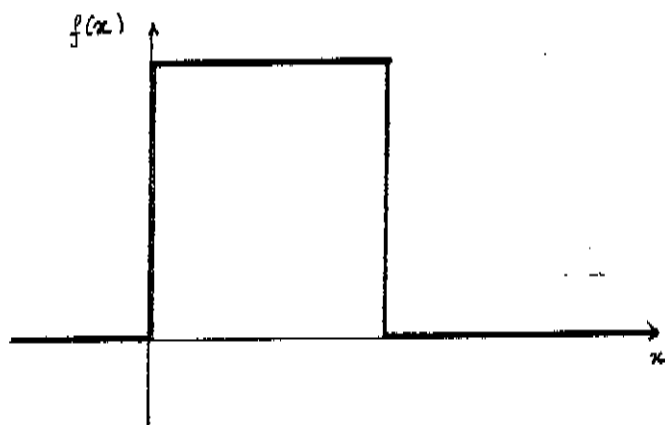
$$G(y) = \iint \dots \int_{g(x) \leq y} f(x) dx
 \tag{5.26}$$

Esempio 7. Somma di variabili casuali

Vogliamo considerare delle variabili casuali indipendenti e trovare la distribuzione della loro somma.

Consideriamo, a titolo di esempio, di avere variabili casuali distribuite uniformemente tra 0 e 1, come possono essere le variabili pseudo-random generate da tipiche routine di calcolatore, e descritte da una densità di probabilità costante

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{altrove} \end{cases} \quad (5.27)$$



Consideriamo una coppia di variabili casuali indipendenti ξ_1, ξ_2 distribuite ognuna secondo la (5.27) e costruiamo la variabile casuale somma

$$\eta = \xi_1 + \xi_2 \quad (5.28)$$

la distribuzione della variabile η sarà descritta dalla probabilità cumulativa

$$\begin{aligned} G(y) &= P\{\xi_1 + \xi_2 \leq y\} = \\ &= \iint_{x_1 + x_2 \leq y} f(x_1) f(x_2) dx_1 dx_2 \end{aligned} \quad (5.29)$$

(54)

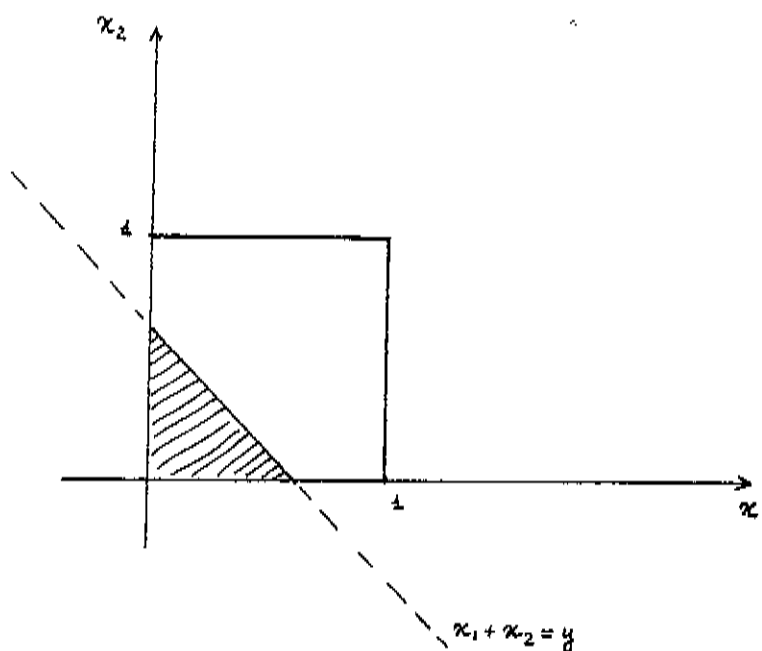
Chiaramente la variabile casuale η sarà distribuita tra 0 e 2, per cui

$$G(y) = 0 \quad \text{se } y < 0$$

$$G(y) = 1 \quad \text{se } y \geq 2$$

Per i valori intermedi, consideriamo (vedi figura) il caso

$$0 < y < 1$$

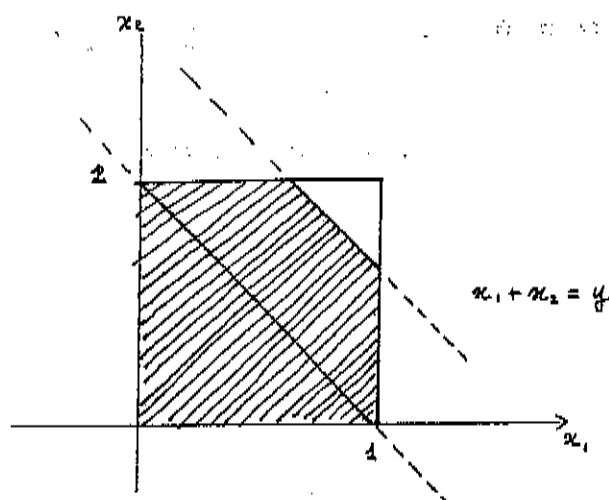


avremo:

$$\begin{aligned} G(y) &= \int_0^y dx_1 \int_0^{y-x_1} dx_2 = \int_0^y dx_1 (y-x_1) dx_1 \\ &= y^2 - \frac{1}{2} y^2 = \frac{1}{2} y^2 \end{aligned}$$

(area del triangolo tratteggiato). Mentre, nel caso

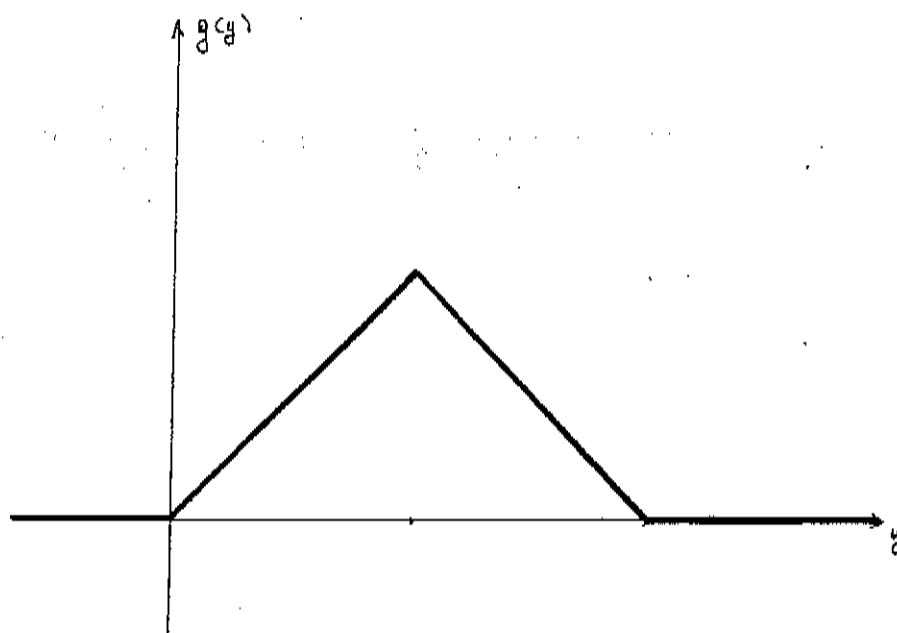
$$1 < y < 2$$



$$\begin{aligned}
 G(y) &= \int_0^{y-1} dx_1 \int_0^1 dx_2 + \int_{y-1}^1 dx_1 \int_0^{y-x_1} dx_2 = \\
 &= (y-1) + \int_{y-1}^1 dx_1 (y-x_1) = \\
 &= y-1 + y(2-y) - \frac{1}{2} (1 - (y-1)^2) = \\
 &= \cancel{y-1} + 2y - y^2 - \cancel{\frac{1}{2}} + \frac{1}{2} y^2 + \cancel{\frac{1}{2}} - \cancel{y} = \\
 &= -\frac{1}{2} y^2 + 2y - 1 \\
 &= 1 - G(2-y)
 \end{aligned}$$

Possiamo descrivere tale distribuzione tramite la densità di probabilità:

$$g(y) = \begin{cases} 0 & y < 0 \\ y & 0 < y < 1 \\ 2-y & 1 < y < 2 \\ 0 & y > 2 \end{cases} \quad (5.30)$$



Considerando ora la somma di 3 variabili casuali

$$\eta = \xi_1 + \xi_2 + \xi_3 \quad (5.31)$$

η risulterà distribuita tra 0 e 3 e dovremo distinguere tra i seguenti casi

$$y < 0 \quad \Rightarrow \quad G(y) = 0$$

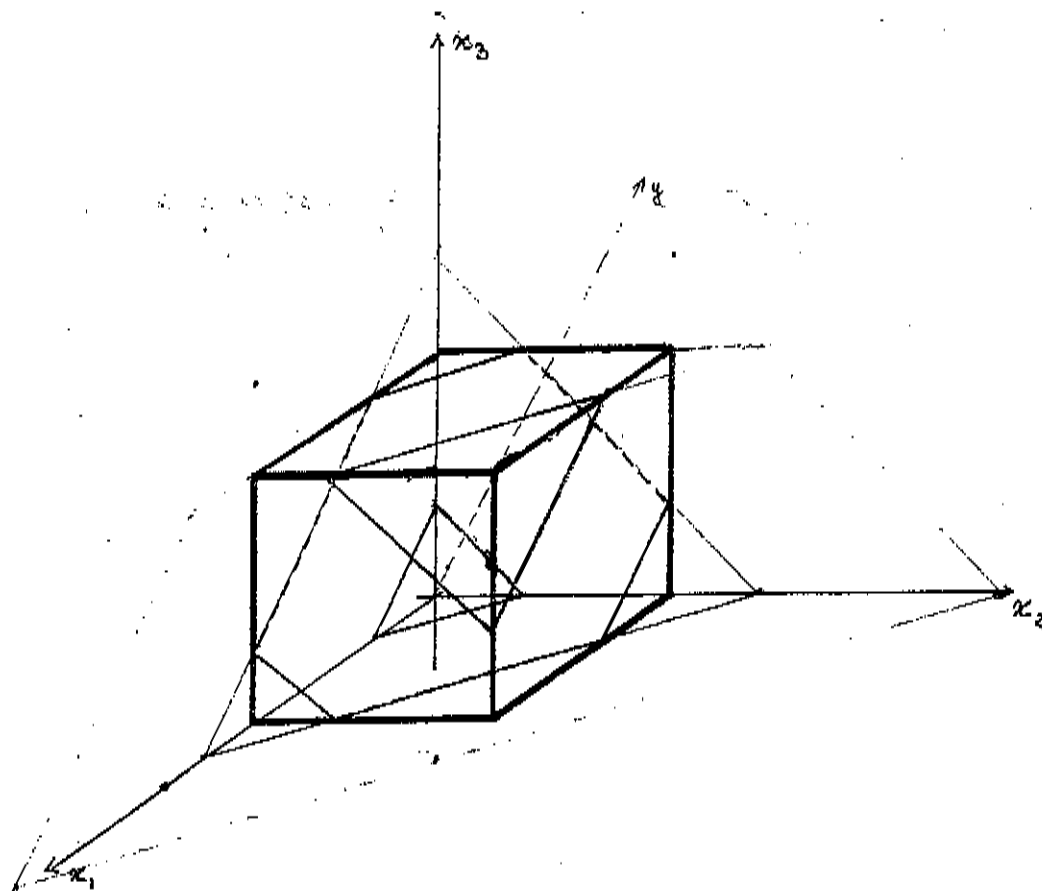
$$y \geq 3 \quad \Rightarrow \quad G(y) = 1$$

$$0 < y < 1 \quad (\text{ogni singolo } \xi_i \text{ sarà minore di } 1)$$

$$1 < y < 2$$

$$2 < y < 3$$

come si può vedere dalla figura, e la cosa, specialmente nel caso intermedio, appare alquanto complicata.

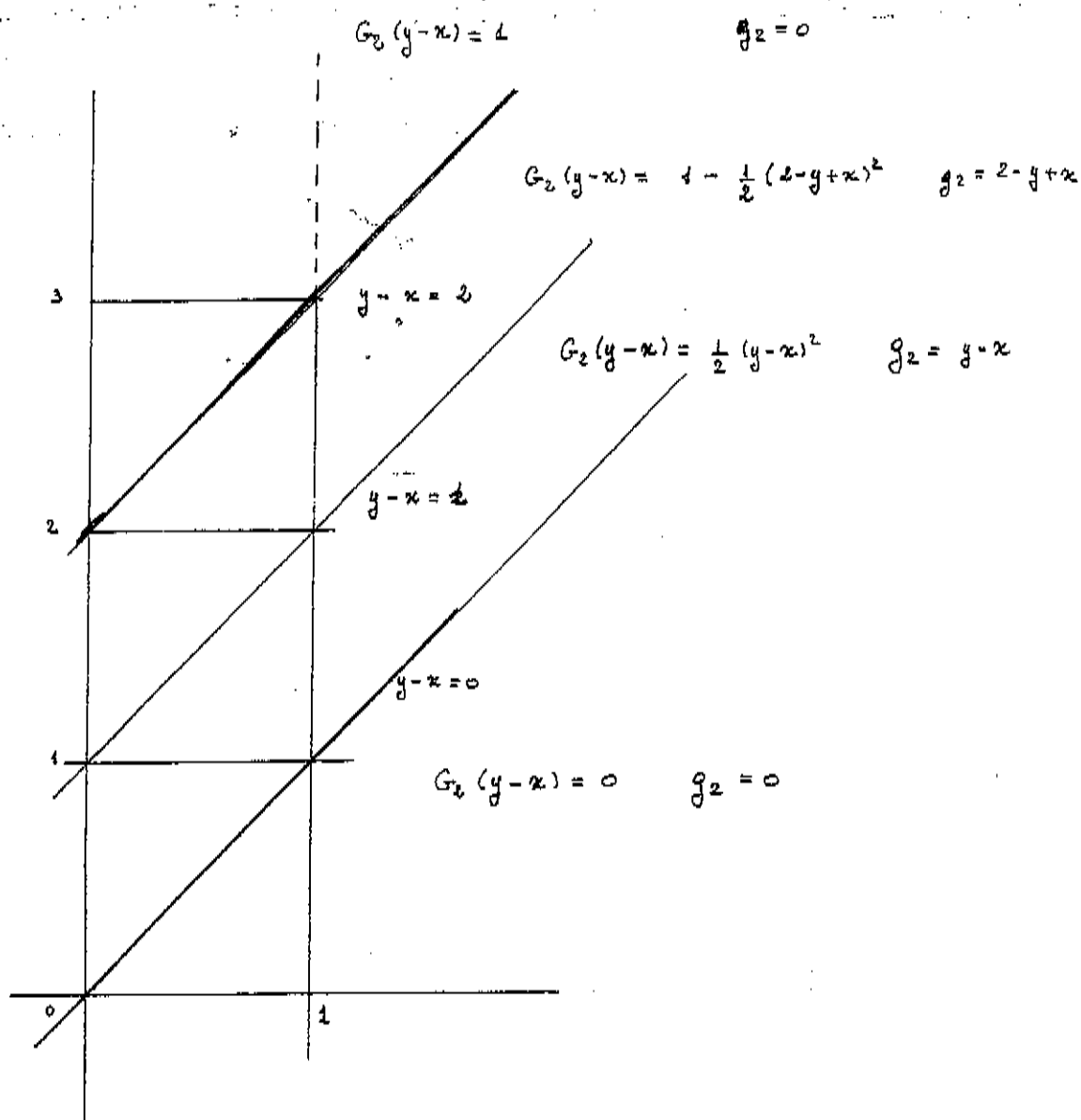


Si possono però sfruttare i risultati già ottenuti. Infatti, detta $G_2(y)$ la probabilità cumulativa del caso precedente e $g_2(y)$ la densità corrispondente (s.30); detta $G_3(y)$ la nuova probabilità cumulativa per tre variabili, abbiamo:

$$\begin{aligned}
 G_3(y) &= \iiint_{x_1+x_2+x_3 \leq y} dx_1 dx_2 dx_3 f(x_1) f(x_2) f(x_3) \\
 &= \int dx_1 \iint_{x_2+x_3 \leq y-x_1} dx_2 dx_3 f(x_1) f(x_2) f(x_3) = \\
 &= \int dx_1 f(x_1) G_2(y-x_1) \\
 &= \int_0^1 dx_1 G_2(y-x_1)
 \end{aligned}$$

$$G_3(y) = \int_0^1 dx G_2(y-x)$$

(5.32)



e, derivando rispetto a y per avere la densità di probabilità

$$g_3(y) = \int_0^1 dx g_2(y-x)$$

(5.33)

allora, risulta chiaro che:

$$y < 0 \Rightarrow g_3(y) = 0$$

$$0 < y < 1 \Rightarrow g_3(y) = \int_0^y dx (y-x) = -\frac{1}{2} (y-x)^2 \Big|_{x=0}^{x=y}$$

$$= \frac{1}{2} y^2$$

$$1 < y < 2 \Rightarrow g_3(y) = \int_0^{y-1} dx (2-y+x) + \int_{y-1}^1 dx (y-x)$$

$$= \frac{1}{2} (2-y+x)^2 \Big|_{x=0}^{x=y-1} - \frac{1}{2} (y-x)^2 \Big|_{x=y-1}^{x=1}$$

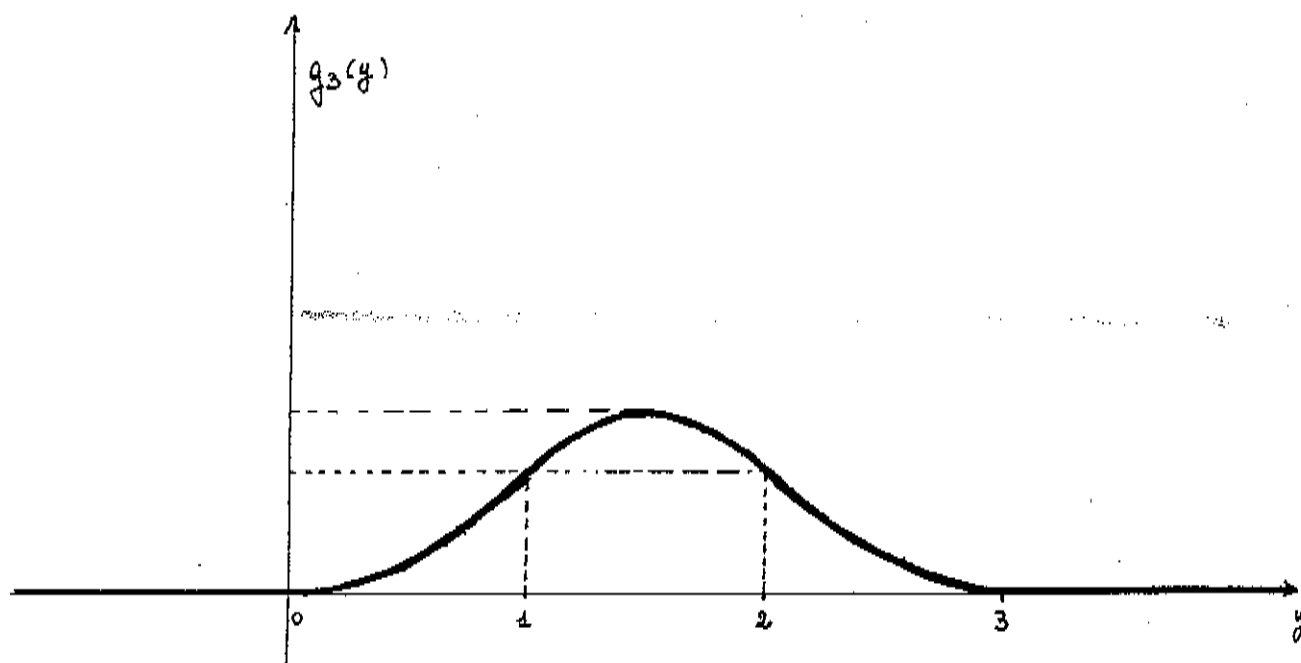
$$= \frac{1}{2} - \frac{1}{2} (2-y)^2 - \frac{1}{2} (y-1)^2 + \frac{1}{2} =$$

$$= 1 - 2 - \frac{1}{2} y^2 + 2y - \frac{1}{2} y^2 - \frac{1}{2} + y =$$

$$= -y^2 + 3y - \frac{3}{2}$$

$$2 < y < 3 \Rightarrow g_3(y) = \int_{y-2}^1 dx (2-y+x) = \frac{1}{2} (2-y+x)^2 \Big|_{x=y-2}^{x=1}$$

$$= \frac{1}{2} (3-y)^2$$



60

Questo procedimento induttivo può essere portato avanti e se consideriamo la somma di n variabili casuali ξ_1, \dots, ξ_n distribuite uniformemente tra 0 e 1

$$\eta = \xi_1 + \xi_2 + \dots + \xi_n \quad (5.34)$$

per la distribuzione di probabilità possiamo scrivere

$$\begin{aligned} G_n(y) &= \int_0^1 dx G_{n-1}(y-x) \\ &= \int_{y-1}^y dx G_{n-1}(x) \end{aligned} \quad (5.35)$$

per quanto riguarda la probabilità cumulativa, mentre, per la densità di probabilità abbiamo:

$$\begin{aligned} g_n(y) &= \int_0^1 dx g_{n-1}(y-x) = \\ &= \int_{y-1}^y dx g_{n-1}(x) = \\ &= G_{n-1}(y) - G_{n-1}(y-1) \end{aligned} \quad (5.36)$$

Notiamo, che per quanto riguarda i valori medi, nel nostro caso le singole variabili casuali ξ_i hanno media

$$\mu = \langle \xi_i \rangle = \int_0^1 x dx = \frac{1}{2}$$

e varianza

$$\sigma^2 = \langle \xi_i^2 \rangle - \frac{1}{4} = \int_0^1 x^2 dx - \frac{1}{4} = \frac{1}{12}$$

mentre la variabile casuale η della (5.34) ha media

$$\langle \eta \rangle = \langle \xi_1 \rangle + \dots + \langle \xi_n \rangle = n\mu = \frac{n}{2}$$

e varianza

$$\sigma_n^2 = n \sigma^2 = \frac{n}{12}$$

Se applichiamo una trasformazione lineare alla variabile casuale η

$$\eta \longrightarrow \xi = \alpha \eta + \beta$$

tramite due parametri α e β possiamo imporre i valori che riteniamo più opportuni per la media

$$\langle \xi \rangle = \alpha \langle \eta \rangle + \beta = \alpha \frac{n}{2} + \beta$$

e per la varianza

$$\sigma_\xi^2 = \alpha^2 \sigma_\eta^2 = \alpha^2 \frac{n}{12}$$

In particolare, scegliendo

$$\begin{cases} \alpha = \sqrt{\frac{12}{n}} \\ \beta = -\sqrt{\frac{12}{n}} \frac{n}{2} = -\sqrt{3n} \end{cases}$$

si ottiene una variabile casuale a media nulla e varianza unitaria

$$\xi = \sqrt{\frac{12}{n}} \left(\xi_1 + \xi_2 + \dots + \xi_n - \frac{n}{2} \right) \quad (5.37)$$

22)

ci si può chiedere cosa succede per $n \rightarrow \infty$, cioè molto grande. La risposta è contenuta nel teorema del limite centrale. Sotto opportune condizioni, verificate dalle variabili del nostro esempio, la variabile (5.37) risulta distribuita secondo una distribuzione gaussiana normalizzata.

Il procedimento di "smussamento" dalla (5.27) alla (5.30), alla (5.33) continua tramite la (5.36) fino a dar luogo ad una distribuzione gaussiana. La cosa notevole è che tale fatto non dipende dalla nostra particolare scelta di variabili casuali distribuite uniformemente, ma il risultato è molto generale e non dipende sostanzialmente dal tipo di distribuzione scelta per le variabili casuali ξ_i .

Da un punto di vista pratico la (5.37) con $n = 12$ (per cui si sommano semplicemente i 12 numeri casuali e si sottrae 6) dà luogo a una densità di probabilità difficilmente distinguibile da una distribuzione gaussiana normalizzata.

$$\xi = \xi_1 + \dots + \xi_{12} - 6$$

$$g_{12}(y) \approx \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

anche se gli effettivi valori di ξ sono compresi tra -6 e 6. Notiamo che per $y = 6$ il valore della gaussiana è $\approx 6 \times 10^{-9}$.

6. Il teorema del limite centrale. Discussione qualitativa

Consideriamo n variabili casuali indipendenti di cui assumiamo siano a media nulla

$$\xi_1, \dots, \xi_n \quad \langle \xi_i \rangle = 0$$

(tale assunzione non è affatto restrittiva) e consideriamo la loro somma

$$\eta = \frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}$$

η risulta ovviamente una variabile casuale a media nulla. Prendiamo in esame la funzione caratteristica per η

$$\begin{aligned} \chi(\lambda) &= \langle e^{-i\lambda\eta} \rangle = \\ &= \langle e^{-i\frac{\lambda}{\sqrt{n}}(\xi_1 + \dots + \xi_n)} \rangle = \\ &= \langle e^{-i\frac{\lambda}{\sqrt{n}}\xi_1} \dots e^{-i\frac{\lambda}{\sqrt{n}}\xi_n} \rangle \end{aligned}$$

essendo queste indipendenti tra loro

$$\begin{aligned} \chi(\lambda) &= \langle e^{-i\frac{\lambda}{\sqrt{n}}\xi_1} \rangle \dots \langle e^{-i\frac{\lambda}{\sqrt{n}}\xi_n} \rangle \\ &= \prod_j \langle e^{-i\frac{\lambda}{\sqrt{n}}\xi_j} \rangle \\ &= e^{\sum_{j=1}^n A_j\left(\frac{\lambda}{\sqrt{n}}\right)} \end{aligned}$$

dove

(64)

$$A_J(\lambda) \equiv \ln \langle e^{-i\lambda \xi_J} \rangle$$

Se espandiamo $A_J\left(\frac{\lambda}{\sqrt{n}}\right)$ per n molto grande:

$$A_J\left(\frac{\lambda}{\sqrt{n}}\right) \approx A_J(0) + \frac{1}{2} A_J''(0) \frac{\lambda^2}{n} + o\left(\frac{\lambda^3}{n^{3/2}}\right)$$

dove il termine di ordine $\frac{\lambda^3}{n^{3/2}}$ è nullo, in quanto

$$A_J'(0) = \frac{d}{d\lambda} \left(\langle e^{-i\lambda \xi_J} \rangle \right)_{\lambda=0} =$$

$$= \frac{1}{\langle e^{-i\lambda \xi_J} \rangle} \left. \langle -i\xi_J e^{-i\lambda \xi_J} \rangle \right|_{\lambda=0}$$

$$= -i \langle \xi_J \rangle = 0$$

Inoltre:

$$A_J''(0) = \left[-\frac{1}{\langle e^{-i\lambda \xi_J} \rangle^2} \langle (-i\xi_J)^2 e^{-i\lambda \xi_J} \rangle + \frac{\langle (-i\xi_J)^2 e^{-i\lambda \xi_J} \rangle}{\langle e^{-i\lambda \xi_J} \rangle} \right]_{\lambda=0}$$

$$= -\langle \xi_J^2 \rangle$$

$$A_J(0) = \ln \langle e^{-i\lambda \xi_J} \rangle \Big|_{\lambda=0} = \ln 1 = 0$$

Pertanto:

$$A_J\left(\frac{\lambda}{\sqrt{n}}\right) \approx -\frac{1}{2} \langle \xi_J^2 \rangle \frac{\lambda^2}{n} + o\left(\frac{\lambda^3}{n^{3/2}}\right)$$

e quindi

$$X(\lambda) \approx e^{-\frac{\lambda^2}{2n} \sum_{j=1}^n \langle \xi_j^2 \rangle + O\left(\frac{1}{\sqrt{n}}\right)}$$

Posto

$$\sigma^2 = \frac{1}{n} \sum_{j=1}^n \langle \xi_j^2 \rangle$$

(che è $O(1)$ e non $O(\frac{1}{n})$) Abbiamo

$$X(\lambda) \approx e^{-\frac{\lambda^2 \sigma^2}{2}}$$

che, ricordando la (4.11), ci esprime la forma di una distribuzione gaussiana con media nulla e varianza σ^2 , pari alla varianza media delle variabili originarie (per $n \rightarrow \infty$ assumiamo che tale media tenda ad un limite finito).

Se al posto di η considerassimo la media aritmetica

$$\bar{\xi} = \frac{1}{n} (\xi_1 + \dots + \xi_n) = \frac{\eta}{\sqrt{n}}$$

allora questa, per n grandi è ~~ancora~~ ancora distribuita come una distribuzione gaussiana (o quasi) con una deviazione standard

$$\frac{\sigma}{\sqrt{n}}$$

questo può essere interpretato dicendo che per n grandi le fluttuazioni statistiche delle medie aritmetiche sono molto basse.

Viceversa se consideriamo la somma semplice delle variabili casuali

66

la fluttuazione ha una larghezza $\sqrt{n}\sigma$, e in molti
casi, per quantità estensive questo ci va bene lo stesso
in quanto \sqrt{n} è piccolo (per grandi n) rispetto a
 n stesso.

Esempio 8.

Supponiamo di avere delle variabili casuali ξ_1, ξ_2, \dots indipendenti tra loro distribuite con la densità di probabilità lorentziana

$$f_i(x) = \frac{a_i}{\pi(a_i^2 + x^2)}$$

dove a_i è una costante positiva. Vogliamo trovare la distribuzione di probabilità della somma

$$\eta = \xi_1 + \xi_2 + \dots + \xi_N$$

Possiamo procedere per induzione. Infatti per la probabilità cumulativa possiamo scrivere

$$G_N(y) = \int \dots \int_{x_1 + \dots + x_N \leq y} f_1(x_1) \dots f_N(x_N) dx_1 \dots dx_N$$

$$= \int dx_N f_N(x_N) G_{N-1}(y - x_N)$$

e per la densità di probabilità:

$$g_N(y) = \int dx g_{N-1}(y - x) f_N(x)$$

con

$$g_1(x) = f_1(x)$$

Abbiamo quindi il seguente tipo di integrale inizialmente

$$\int \frac{a_1}{\pi(a_1^2 + (y-x)^2)} \frac{a_2}{\pi(a_2^2 + x^2)} dx =$$

$$\begin{aligned}
&= \frac{a_1 a_2}{\pi^2} \int_{-\infty}^{+\infty} \frac{1}{a_1^2 + (y-x)^2} \frac{1}{a_2^2 + x^2} dx \\
&= \frac{a_1 a_2}{\pi^2} \cdot 2\pi i \left\{ \begin{aligned} &\text{Res}_{z=y+ia_1} \frac{1}{a_1^2 + (y-z)^2} \cdot \frac{1}{a_2^2 + z^2} + \\ &\text{Res}_{z=ia_2} \frac{1}{a_1^2 + (y-z)^2} \cdot \frac{1}{a_2^2 + z^2} \end{aligned} \right\} \\
&= \frac{2 a_1 a_2}{\pi} i \left\{ \frac{1}{2ia_1} \frac{1}{a_2^2 + (y+ia_1)^2} + \frac{1}{a_1^2 + (y-ia_2)^2} \frac{1}{2ia_2} \right\} \\
&= \frac{a_1 a_2}{\pi} \left\{ \frac{1}{a_1} \frac{1}{(y+ia_1+ia_2)(y+ia_1-ia_2)} + \frac{1}{a_2} \frac{1}{(y+ia_1-ia_2)(y-ia_1-ia_2)} \right\} \\
&= \frac{a_1 a_2}{\pi} \frac{1}{y+ia_1-ia_2} \frac{1}{a_1 a_2 (y+ia_1+ia_2)(y-ia_1-ia_2)} \\
&= \frac{1}{\pi} \frac{1}{y+ia_1-ia_2} \frac{a_2(y-ia_1-ia_2) + (y+ia_1+ia_2)a_1}{y^2 + (a_1+a_2)^2} \\
&= \frac{1}{\pi} \frac{1}{(a_1+a_2)^2 + y^2} \frac{y(a_1+a_2) - i(a_1 a_2 + a_2^2) + i(a_1^2 + a_1 a_2)}{y + i(a_1 - a_2)} \\
&= \frac{1}{\pi} \frac{1}{(a_1+a_2)^2 + y^2} \frac{y(a_1+a_2) + i(a_1+a_2)(a_1-a_2)}{y + i(a_1 - a_2)} \\
&= \frac{1}{\pi} \frac{1}{(a_1+a_2)^2 + y^2} \frac{(a_1+a_2)(y + i(a_1-a_2))}{y + i(a_1-a_2)} = \\
&= \frac{(a_1+a_2)}{\pi} \frac{1}{(a_1+a_2)^2 + y^2}
\end{aligned}$$

Logo

$$\int_{-\infty}^{+\infty} \frac{a_1}{\pi(a_1^2 + (y-x)^2)} \frac{a_2}{\pi(a_2^2 + x^2)} dx = \frac{a_1 + a_2}{\pi((a_1+a_2)^2 + y^2)}$$

Portando, iterando tale relazione.

$$g_1(x) = \frac{a_1}{\pi (a_1^2 + x^2)}$$

$$g_2(x) = \frac{a_1 + a_2}{\pi [(a_1 + a_2)^2 + x^2]}$$

$$g_3(x) = \frac{a_1 + a_2 + a_3}{\pi [(a_1 + a_2 + a_3)^2 + x^2]}$$

⋮

$$g_N(x) = \frac{a_1 + \dots + a_N}{\pi [(a_1 + \dots + a_N)^2 + x^2]}$$

Considerando la media

$$\frac{1}{N} \sum_{j=1}^N a_j = a^{(N)}$$

abbiamo

$$g_N(x) = \frac{N a^{(N)}}{\pi [(N a^{(N)})^2 + x^2]}$$

Considerando la variabile casuale

$$\xi = \frac{1}{\sqrt{N}} (\xi_1 + \dots + \xi_N) = \frac{\eta}{\sqrt{N}}$$

avremo la sua probabilità cumulativa

$$F_N(y) = \mathcal{P}\{\xi \leq y\} = \mathcal{P}\{\eta \leq \sqrt{N}y\} = G_N(\sqrt{N}y)$$

e la densità di probabilità:

$$f_N(x) = \sqrt{N} g_N(\sqrt{N} x)$$

$$f_N(x) = \frac{1}{\sqrt{N}} \frac{N^2 a^{(N)}}{\pi [N^2 (a^{(N)})^2 + N x^2]} = \frac{\sqrt{N} a^{(N)}}{\pi [(\sqrt{N} a^{(N)})^2 + x^2]}$$

Se ad esempio assumiamo tutti gli a_i uguali

$$f_N(x) = \frac{\sqrt{N} a}{\pi [(\sqrt{N} a)^2 + x^2]}$$

Se invece consideriamo i valori medie aritmetiche

$$\bar{x} = \frac{1}{N} (\xi_1 + \dots + \xi_N)$$

abbiamo

$$f_N^{(m)}(x) = \frac{a}{\pi [a^2 + x^2]}$$

e tutto questo contrasta con il teorema del limite centrale.

Il motivo risiede nel fatto che la distribuzione lorentziana

$$f(x) = \frac{a}{\pi (a^2 + x^2)}$$

pur essendo una valida densità di probabilità, è sostanzialmente priva di momenti. Ad es:

$$\langle \xi^2 \rangle = \int_{-\infty}^{+\infty} \frac{a x^2 dx}{\pi (a^2 + x^2)} = \infty$$

7. Momenti e cumulanti.

Abbiamo visto che la definizione di varianza per una variabile casuale ξ è

$$\sigma^2 = \langle \xi^2 \rangle - \langle \xi \rangle^2 \quad (7.1)$$

mentre nel caso di due variabili ξ_i, ξ_j una misura della loro correlazione può essere definita dalla differenza

$$\rho_{ij} = \langle \xi_i \xi_j \rangle - \langle \xi_i \rangle \langle \xi_j \rangle \quad (7.2)$$

per cui a volte σ^2 viene chiamato secondo valore di correlazione e la (7.2) riproduce la (7.1) quando $i=j$. Questi vengono detti anche con altri nomi, cumulanti (di ordine 2), e si usa la simbologia

$$\langle \xi^2 \rangle_c = \langle \xi^2 \rangle - \langle \xi \rangle^2 \quad (7.3)$$

Per variabili differenti ($i \neq j$) il cumulante $\langle \xi_i \xi_j \rangle_c$ è definito dalla relazione

$$\langle \xi_i \xi_j \rangle = \langle \xi_i \rangle \langle \xi_j \rangle + \langle \xi_i \xi_j \rangle_c \quad (7.4)$$

Il cumulante $\langle \xi_i \xi_j \rangle_c$ contiene le stesse informazioni del momento $\langle \xi_i \xi_j \rangle$ dello stesso ordine 2, ma spesso in maniera più evidente. Ad esempio se ξ_i e ξ_j sono variabili indipendenti allora il cumulante $\langle \xi_i \xi_j \rangle_c$ è nullo. L'equazione (7.3) è un caso particolare della (7.4), quando $i=j$.

Il concetto di cumulante viene generalizzato anche ai momenti ulteriori di grado più elevato. Ad esempio, per tre variabili casuali abbiamo la definizione:

$$\begin{aligned}
 \langle \xi_i \xi_j \xi_k \rangle &= \langle \xi_i \rangle \langle \xi_j \rangle \langle \xi_k \rangle \\
 &+ \langle \xi_i \xi_j \rangle_c \langle \xi_k \rangle + \langle \xi_i \xi_k \rangle_c \langle \xi_j \rangle + \langle \xi_j \xi_k \rangle_c \langle \xi_i \rangle \\
 &+ \langle \xi_i \xi_j \xi_k \rangle_c
 \end{aligned} \tag{7.5}$$

che definisce appunto il cumulante $\langle \xi_i \xi_j \xi_k \rangle_c$.

L'idea di base è che il cumulante contiene informazioni sulla correlazione tra le varie variabili. Se abbiamo tre variabili la correlazione può essere dovuta a correlazioni di coppia (le cui informazioni sono contenute nei cumulanti di ordine due $\langle \xi_i \xi_j \rangle_c$, $\langle \xi_i \xi_k \rangle_c$, $\langle \xi_j \xi_k \rangle_c$) e, a parte ciò, da una correlazione "pura" tra le tre. Ad esempio, può capitare che considerate a coppie le tre variabili sono mutualmente indipendenti, per cui $\langle \xi_i \xi_j \rangle_c = 0$, $\langle \xi_i \xi_k \rangle_c = 0$, $\langle \xi_j \xi_k \rangle_c = 0$, mentre $\langle \xi_i \xi_j \xi_k \rangle_c$ è non nullo e le tre variabili non sono indipendenti considerate come terz.

Nel caso $i=j=k$ la (7.5) definisce il cumulante $\langle \xi^3 \rangle_c$ come:

$$\langle \xi^3 \rangle = \langle \xi \rangle^3 + 3 \langle \xi^2 \rangle_c \langle \xi \rangle + \langle \xi^3 \rangle_c \tag{7.6}$$

Possiamo procedere all'ordine 4:

$$\begin{aligned}
 \langle \xi_i \xi_j \xi_k \xi_l \rangle &= \langle \xi_i \rangle \langle \xi_j \rangle \langle \xi_k \rangle \langle \xi_l \rangle \\
 &+ \langle \xi_i \xi_j \rangle_c \langle \xi_k \rangle \langle \xi_l \rangle + \dots \quad (6 \text{ termini} = \binom{4}{2} = \frac{4!}{2!1!1!}) \\
 &+ \langle \xi_i \xi_j \rangle_c \langle \xi_k \xi_l \rangle_c + \dots \quad (3 \text{ termini}) \\
 &+ \langle \xi_i \xi_j \xi_k \rangle_c \langle \xi_l \rangle + \dots \quad (4 \text{ termini}) \\
 &+ \langle \xi_i \xi_j \xi_k \xi_l \rangle_c
 \end{aligned} \tag{7.7}$$

e nel caso particolare di una sola variabile

$$\langle \xi^4 \rangle = \langle \xi \rangle^4 + 6 \langle \xi^2 \rangle_c \langle \xi \rangle^2 + 3 \langle \xi^2 \rangle_c^2 + 4 \langle \xi^3 \rangle_c \langle \xi \rangle + \langle \xi^4 \rangle_c \tag{7.8}$$

⑤

1

7.9

3.10

(4)

e così via

Notiamo quindi che ogni cumulante può essere espresso tramite i valori medi, cioè i momenti fino all'ordine stesso del cumulante.

Abbiamo anche l'ulteriore proprietà che l'operazione di calcolo del cumulante si mantiene inalterata per trasformazioni lineari; cioè, ad esempio se

$$\eta_i = \sum_j \alpha_{ij} \xi_j$$

con α_{ij} costanti della trasformazione lineare di variabili casuali; allora

$$\langle \eta_i \eta_j \rangle_c = \sum_{k, l} \alpha_{ik} \alpha_{jl} \langle \xi_k \xi_l \rangle_c$$

e lo stesso vale per prodotti di ordine più elevato.

Notiamo però che ciò non è vero per trasformazioni non lineari. Ad esempio, assumiamo

$$\langle \xi \rangle = 0 \quad \eta = \xi^2$$

allora

$$\begin{aligned} \langle \eta^2 \rangle_c &= \langle \eta^2 \rangle - \langle \eta \rangle^2 = \\ &= \langle \xi^4 \rangle - \langle \xi^2 \rangle^2 \end{aligned}$$

$$\langle \xi^4 \rangle_c = \langle \xi^4 \rangle - 3 \langle \xi^2 \rangle^2$$

per cui

$$\langle \eta^2 \rangle_c \neq \langle \xi^4 \rangle_c$$

8. Teorema di espansione dei cumulanti.

Questo teorema stabilisce che, definita la relazione

$$\langle 1 \rangle_c = 0$$

(il cumulante di una costante è definito nullo), allora:

$$\langle e^{-i\lambda \xi} \rangle = e^{\langle e^{-i\lambda \xi} \rangle_c}$$

con

$$\langle e^{-i\lambda \xi} \rangle_c = \sum_{n=1}^{\infty} \frac{(-i\lambda)^n}{n!} \langle \xi^n \rangle_c$$

La funzione $\langle e^{-i\lambda \xi} \rangle_c$ può essere chiamata la funzione caratteristica cumulante.

Per provare questo teorema procediamo in questo modo:
Consideriamo la funzione caratteristica

$$\chi(\lambda) = \langle e^{-i\lambda \xi} \rangle = \sum_{m=0}^{\infty} \frac{(-i\lambda)^m}{m!} \langle \xi^m \rangle$$

e scriviamo $\langle \xi^m \rangle$ come combinazione di cumulanti procedendo in maniera grafica. Disegniamo m punti e dividiamoli in gruppi in tutti i modi possibili, connettendo fra loro i punti interni a un gruppo.

Avremo così m_1 gruppi con un punto solo, m_2 con due punti, ..., m_n gruppi con n punti, e così via, con

$$m = \sum_n n m_n$$

($m_n \geq 0$). Il modo in cui otteniamo un raggruppamento con $m_1, m_2, \dots, m_n, \dots$ fissati è dato da

$$\frac{m!}{m_1! (1!)^{m_1} m_2! (2!)^{m_2} \dots m_n! (n!)^{m_n} \dots}$$

Infatti: gli m punti hanno $m!$ permutazioni. Lo scambio di punti all'interno di un gruppo non produce alcun cambiamento, per cui dividiamo per $n!$ per ogni gruppo di n punti (da cui il fattore $(n!)^{m_n}$). Lo scambio di gruppi con lo stesso numero di punti ugualmente non produce alcun cambiamento, da cui il fattore $m_n!$.

Possiamo allora scrivere

$$\begin{aligned} \langle \xi^m \rangle &= \sum_{m_1, \dots, m_n, \dots} m! \frac{1}{m_1!} \left(\frac{\langle \xi \rangle}{1!} \right)^{m_1} \dots \frac{1}{m_n!} \left(\frac{\langle \xi^n \rangle_c}{n!} \right)^{m_n} \dots \\ &\quad \sum n m_n = m \\ &= \sum_{m_1, \dots, m_n, \dots} m! \prod_n \frac{1}{m_n!} \left(\frac{\langle \xi^n \rangle_c}{n!} \right)^{m_n} \delta \left(m - \sum_n n m_n \right) \end{aligned}$$

e quindi

$$\begin{aligned} \chi(\lambda) &= \sum_{m_1, \dots, m_n, \dots} (-i\lambda)^m \prod_n \frac{1}{m_n!} \left(\frac{\langle \xi^n \rangle_c}{n!} \right)^{m_n} \delta \left(m - \sum_n n m_n \right) \\ &= \sum_{m_1, \dots, m_n, \dots} \prod_n \frac{1}{m_n!} \left(\frac{(-i\lambda)^n \langle \xi^n \rangle_c}{n!} \right)^{m_n} \\ &= \prod_{n=1}^{\infty} \sum_{m_n=0}^{\infty} \frac{1}{m_n!} \left(\frac{(-i\lambda)^n \langle \xi^n \rangle_c}{n!} \right)^{m_n} \\ &= \prod_{n=1}^{\infty} e^{\frac{(-i\lambda)^n \langle \xi^n \rangle_c}{n!}} \\ &= e^{\sum_{n=1}^{\infty} \frac{(-i\lambda)^n \langle \xi^n \rangle_c}{n!}} \\ &= e^{\langle e^{-i\lambda \xi} \rangle_c} = \chi_c(\lambda) \end{aligned}$$

Da cui il risultato

3. Teorema del limite centrale (esposizione generale)

Abbiamo visto che la funzione caratteristica di una distribuzione gaussiana con media nulla ($\mu = 0$) e varianza σ^2 è data da:

$$\chi(\lambda) = e^{-\frac{1}{2}\sigma^2\lambda^2}$$

da cui

$$\chi_c(\lambda) = -\frac{1}{2}\sigma^2\lambda^2$$

e quindi:

$$\langle \xi^2 \rangle_c = \sigma^2$$

$$\langle \xi^n \rangle_c = 0 \quad n \neq 2$$

In queste proprietà è contenuta tutta la distribuzione gaussiana. Vediamo ora di stabilire con chiarezza le condizioni di validità del teorema del limite centrale.

Siano $\xi_1, \dots, \xi_N, \dots$ variabili casuali che soddisfino le seguenti condizioni:

a) $\langle \xi_i \rangle = 0$

b) $\frac{1}{N} \sum_{i,j} \langle \xi_i \xi_j \rangle_c \equiv \sigma^2 = O(1)$

c) $\frac{1}{N} \sum_{i,j,k} \langle \xi_i \xi_j \xi_k \rangle_c \equiv \mu_3 \leq O(1)$

$$\frac{1}{N} \sum_{j_1, \dots, j_n} \langle \xi_{j_1} \dots \xi_{j_n} \rangle_c \equiv \mu_n \leq O(N^{(n-3)/2})$$

Allora, posto

$$\eta = \frac{1}{\sqrt{N}} (\xi_1 + \dots + \xi_N)$$

abbiamo

$$i) \quad \langle \eta \rangle = 0$$

$$ii) \quad \langle \eta^2 \rangle_c = \sigma^2$$

$$iii) \quad \langle \eta^n \rangle_c \leq O\left(\frac{1}{\sqrt{N}}\right) \quad n > 2$$

I primi due risultati sono ovvi. Per il risultato (iii) notiamo che

$$\langle \eta^3 \rangle_c = \left(\frac{1}{N}\right)^{3/2} \sum_{i,j,k} \langle \xi_i \xi_j \xi_k \rangle_c = \frac{\mu_3}{\sqrt{N}}$$

e analogamente

$$\langle \eta^n \rangle_c = \left(\frac{1}{N}\right)^{n/2} \sum_{j_1, \dots, j_n} \langle \xi_{j_1} \dots \xi_{j_n} \rangle_c = \frac{\mu_n}{\sqrt{N}^{n-2}}$$

$$\leq O\left(\sqrt{\frac{N^{n-3}}{N^{n-2}}}\right) = O\left(\frac{1}{\sqrt{N}}\right)$$

Vediamo ora di chiarire il significato di tale teorema.

- (1). La conclusione (iii) dice che a parte termini dell'ordine di $\frac{1}{\sqrt{N}}$ i valori medi di η possono essere calcolati tramite una distribuzione gaussiana. Abbiamo però una ovvia restrizione: i momenti più elevati possono risultare errati perché $\langle \eta^n \rangle$ può contenere molti termini quando sono espressi mediante i cumulanti. Se il numero di termini si avvicina a \sqrt{N} allora i termini dell'ordine di $\frac{1}{\sqrt{N}}$ non possono

essere più trascurati. Pertanto la distribuzione della variabile ~~casuale~~ η si può considerare gaussiana se ci limitiamo al calcolo delle potenze più basse di η

(2). Se $\xi_1, \dots, \xi_N, \dots$ sono variabili mutuamente indipendenti allora le relazioni (b) e (c) sono verificate in quanto σ_2, μ_n sono tutte $O(1)$

(3). Se non sono indipendenti ma ogni variabile è correlata con al massimo m altre con $m \ll N$ allora le somme (c) coinvolgono al massimo $N m^n$ termini e tutto è ok.

(4.) Da un punto di vista pratico si calcolano in genere solo le potenze medie più basse di η e non è necessario verificare l'indipendenza delle variabili.